

AD-A217 607

AD-E50118L
Copy 24 of 150 copies

(2)

IDA PAPER P-2310

SUMMARY REPORT OF THE
DEFENSE SCIENCE STUDY GROUP
1985 - 1988

DTIC
ELECTE
JAN 31 1990
S D

Richard J. Bergemann
Nancy P. Licato

September 1989

Prepared for
Defense Advanced Research Projects Agency

DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited



INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street, Alexandria, Virginia 22311-1772

90 01 31 086

IDA Log No. HQ 89-34853

DEFINITIONS

IDA publishes the following documents to report the results of its work.

Reports

Reports are the most authoritative and most carefully considered products IDA publishes. They normally embody results of major projects which (a) have a direct bearing on decisions affecting major programs, (b) address issues of significant concern to the Executive Branch, the Congress and/or the public, or (c) address issues that have significant economic implications. IDA Reports are reviewed by outside panels of experts to ensure their high quality and relevance to the problems studied, and they are released by the President of IDA.

Group Reports

Group Reports record the findings and results of IDA established working groups and panels composed of senior individuals addressing major issues which otherwise would be the subject of an IDA Report. IDA Group Reports are reviewed by the senior individuals responsible for the project and others as selected by IDA to ensure their high quality and relevance to the problems studied, and are released by the President of IDA.

Papers

Papers, also authoritative and carefully considered products of IDA, address studies that are narrower in scope than those covered in Reports. IDA Papers are reviewed to ensure that they meet the high standards expected of refereed papers in professional journals or formal Agency reports.

Documents

IDA Documents are used for the convenience of the sponsors or the analysts (a) to record substantive work done in quick reaction studies, (b) to record the proceedings of conferences and meetings, (c) to make available preliminary and tentative results of analyses, (d) to record data developed in the course of an investigation, or (e) to forward information that is essentially unanalyzed and unevaluated. The review of IDA Documents is suited to their content and intended use.

The work reported in this document was conducted under contract MDA 903 84 C 0031 for the Department of Defense. The publication of this IDA Paper does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that Agency.

This Paper has been reviewed by IDA to assure that it meets the high standards of thoroughness, objectivity, and appropriate analytical methodology and that the results, conclusions and recommendations are properly supported by the material presented.

Approved for public release; distribution unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE September 1989	3. REPORT TYPE AND DATES COVERED Final--October 1985 - November 1988
4. TITLE AND SUBTITLE Summary Report of the Defense Science Study Group		5. FUNDING NUMBERS C - MDA 903 84 C 0031 T - A-103	
6. AUTHOR(S) Richard J. Bergemann, Nancy P. Licato		8. PERFORMING ORGANIZATION REPORT NUMBER IDA Paper P-2310	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Defense Analyses 1801 N. Beauregard St. Alexandria, VA 22311		10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22209-2308		11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.		12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The Defense Science Study Group is a DARPA sponsored program that identifies some of the most talented young scientists and engineers in the country today and exposes them to the outstanding issues of national defense. The program seeks to strengthen ties with the part of the scientific and engineering community that is outside of the defense establishment. Its goals are to provide an education to this select group of individuals on the broad scope of critical defense-related scientific and technical problems, encourage long-term, personal involvement in their solution, and provide DoD with a new source of technical advisors and informed critics. The program resides at IDA and is guided by the advice of mentors who have had distinguished careers in defense or academia. Individuals spend about twenty days a year for three years listening to presentations given by distinguished speakers, visiting defense facilities and conducting defense studies on topics of interest. The report summarizes the program's first three years of activities from 1985-88.			
14. SUBJECT TERMS Defense Science Study Group, defense technology, defense-related academic research, defense science research		15. NUMBER OF PAGES 148 16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR

IDA PAPER P-2310

SUMMARY REPORT OF THE
DEFENSE SCIENCE STUDY GROUP
1985 - 1988

Richard J. Bergemann
Nancy P. Licato



September 1989

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	For
A-1	



INSTITUTE FOR DEFENSE ANALYSES

Contract MDA 903 84 C 0031
DARPA Assignment A-103

ABSTRACT

The Defense Science Study Group is a DARPA sponsored program that identifies some of the most talented young scientists and engineers in the country today and exposes them to the outstanding issues of national defense. The program seeks to strengthen ties with the part of the scientific and engineering community that is outside of the defense establishment. Its goals are to provide an education to this select group of individuals on the broad scope of critical defense-related scientific and technical problems, encourage long-term, personal involvement in their solution, and provide DoD with a new source of technical advisors and informed critics. The program resides at IDA and is guided by the advice of mentors who have had distinguished careers in defense or academia. Individuals spend about twenty days a year for three years listening to presentations given by distinguished speakers, visiting defense facilities and conducting defense studies on topics of interest. This report summarizes the program's first three years of activities from 1985-88.

CONTENTS

Abstract	iii
I. INTRODUCTION.....	I-1
A. Objective	I-1
B. Program Characteristics	I-2
II. EDUCATIONAL PROGRAM.....	II-1
A. Initial Year	II-2
B. Second Year	II-5
C. Final Year	II-7
III. STUDIES AND ANALYSES.....	III-1
A. Monitoring Yields of Underground Nuclear Tests Using Hydrodynamic Methods.....	III-3
B. Opportunities for Long-Wavelength IR Detectors.....	III-45
C. Uses of Charged Particle Beams for Transportable HF Radar	III-85
D. Popeye--An Expendable Battlefield Surveillance System	III-103
E. Some Promising Directions in Applied Mathematics	III-117
IV. CONCLUSION.....	IV-1
APPENDIX A--Participants and Mentors During the Period	A-1
APPENDIX B--Some Distinguished Speakers	B-1

I. INTRODUCTION

Today's complex technical issues associated with national defense require the attention of the best scientists and engineers in the country. From World War II until the Vietnam conflict, a close link existed between the elite of the scientific community and the defense establishment. This link not only helped ensure that the nation's defense needs were met, but also provided knowledgeable technical criticism of the highest quality. This link was weakened substantially during the Vietnam conflict, thus depriving the defense establishment of access to many of the country's most talented researchers for both contributions and informed criticism.

A long-standing strength of the Institute for Defense Analyses (IDA) has been its ability to provide an atmosphere in which the scientific community could become aware of the specific technical content of national security issues and in which scientists could carry out academic research on defense-related technical problems. Recognizing this, in FY 1986 the Defense Advanced Research Projects Agency (DARPA) established the Defense Science Study Group (DSSG) at IDA to identify a select group of young scientists and engineers in the country who were mostly outside the defense community. They were to be provided with an education on national defense and its associated technical problems.

A. OBJECTIVE

The principal objectives of the DSSG are to reestablish and strengthen links between the Department of Defense and the scientific and engineering communities and expand the quality of technical advisors available to the Department of Defense. These are accomplished in a program that fosters an interest among some of the country's brightest young scientists in the technical aspects of national defense issues combining education on a broad range of defense topics with independent studies and analyses of technical defense problems of interest. Program participants acquire an understanding of the comprehensive nature of national defense activities and the difficult technical challenges that need to be met. It is hoped that participants will eventually provide new insights on defense problems, guide some of the most promising students of today into defense careers, and play an active role in the defense community in the future. The program seeks to foster a

more complete understanding of the broader issues associated with the defense of this country among those individuals who are likely to be among the most influential and respected members of the scientific and engineering community of tomorrow.

B. PROGRAM CHARACTERISTICS

The DSSG is characterized by its multidisciplinary nature and the rigorous and careful process that is followed to select participants. The names of candidates can be suggested by participants and mentors or solicited from organizations such as the National Science Foundation, the Office of Science and Technology Policy, and the Sloan Foundation. Nominations are also taken from outstanding individuals in any field of science or technology. The selection of candidates invited to join is made by IDA after consulting with a variety of senior individuals for references and after the candidate has been approached to establish interest in the program. The maximum length of time that participants can remain in the program is three years. Participants are selected primarily from the academic community and are generally in the first decade of their career. A list of the 17 individuals who were members during this period of time can be found in Appendix A.

A group of senior mentors who have had distinguished careers in defense or academia serve as advisors. They help identify candidates, suggest defense problems of importance to study, advise IDA on the conduct of the program, independently review the technical work accomplished, and assess the overall success of the program. A list of the mentors who have served during this period can also be found in Appendix A.

The Institute for Defense Analyses directs and administers the DSSG program and gives it continuity. It selects all participants, organizes the program's agenda, arranges all program activities, and provides the necessary administrative support. In addition, by virtue of its own active defense research and analysis program, it provides a convenient source of in-house expertise on a variety of defense topics. IDA is also responsible for fostering the development of links between members of the defense community and DSSG participants.

DARPA, as the program sponsor, provides overall guidance to the program, assists in developing the program's technical agenda, and most directly benefits from the results of the DSSG's activities.

DSSG activities are split between education and studies and analyses during the approximately 20 days each year that participants devote to the program. The educational portion of the program is structured around a number of major national security issues. The introduction to these topics includes presentations by defense experts and organized travel to major defense facilities.

Research studies and analyses are conducted by the participants and represent an extension of the educational process whereby participants learn about the process of performing defense research and develop links to the defense community. Participants choose their own topics within very broad guidelines and work alone or organize themselves into groups as they see fit. They are provided access to both classified and unclassified resources through IDA. Participants are encouraged to work on research projects when the DSSG is not formally in session and the IDA facilities and services are available to the participants upon request.

II. EDUCATIONAL PROGRAM

A formal introduction into the defense community was provided by a comprehensive program of lectures and visits to defense facilities that dealt with all aspects of national defense. Figure 1 summarizes the topics that were covered from 1985-88. While the focus of the program was on science and technology, much time was spent developing the military context that gives issues in science and technology their importance in the national defense arena.

	DEFENSE	AIR FORCE	NAVY	ARMY
Military Operations	<ul style="list-style-type: none"> - National C³ (Pentagon, WH staff) (1) 	<ul style="list-style-type: none"> - Strategic Nuclear Forces-Air-C³ (SAC) (1) - Surveillance--C³ (Cheyenne Mtn.--NORAD, satellite systems) (1) 	<ul style="list-style-type: none"> Command & Control--CINCPAC and CINCPACFLT (2) - Surveillance--(2) - Training (2) - Strategic Nuclear Forces--Sea (2) - ASW (2) - Special Forces (SEALS) (2) 	<ul style="list-style-type: none"> - Training--Combined Arms (3)
Science & Technology	<ul style="list-style-type: none"> - IR detector Production (1) - Superconductivity (2) - Nuclear Effects (1) - Strategic Surveillance Sensors and Technologies (1) - Nuclear Explosions Detection Underground (1) - DARPA Strategic Computing Program (1) - Image Processing (1) - Missile Plume Phenomenology (1) - Robotics (1) - Nuclear Weapons Testing (3) - Isotope Separation (3) - Free Electron Laser (3) 	<ul style="list-style-type: none"> - Aerospace Plane (computational fluid dynamics, materials combustion, support) (1) - MILSTAR (1) 	<ul style="list-style-type: none"> - Marine Mammal Research (2) - ASW (2) - Microelectronics (2) - Submarine Laser Communications (2) - Arctic Ice Research (2) - Torpedoes (2) - Submarine (2) - Advanced Surveillance (2) - NAVSTAR/GPS (1) 	<ul style="list-style-type: none"> - Night Vision (1)
Policy	<ul style="list-style-type: none"> - Executive Branch (OMB, WH) (2) - Congress (2) - Arms Control (1) 		<ul style="list-style-type: none"> - Maritime Strategy (2) 	
Non-Defense	<ul style="list-style-type: none"> - Space Debris (2) - Anti-Terrorism (2) - Manned Space Flight (3) - Controlled Nuclear Fusion (NOVA Laser) (3) 			

(-) = Year

Figure 1. Topics Covered in Presentations and Site Visits

Many distinguished speakers addressed the DSSG from 1985-88 on a wide range of topics. A list of some of them can be found in Appendix B. A year-by-year account of the lectures and visits to defense facilities follows. Also included is a brief overview of the sessions devoted to research studies and analyses.

A. INITIAL YEAR

In its first year of operation, the DSSG convened five times, as shown in Figure 2. The activities pursued during this year are summarized in Figure 3. Participants for the year totaled 12. Fifteen mentors joined the program. Detailed information on the DSSG activities for 1986 can be found in IDA Memorandum Report M-309, May 1987, and is summarized by session as follows.

Introductory Session	December 8-9, 1985
Overview Session	April 19-21, 1986
First Working Session	July 14-24, 1986
Second Working Session	September 3-12, 1986
Review Session	November 7-8, 1986

Figure 2. Schedule for 1986

1. Introductory Session

The purpose of the first meeting of the DSSG was for the participants to receive an introduction to defense technology problems and meet each other, the mentors, IDA staff responsible for the program and DARPA staff. Dr. James Tegnalia, Deputy Director of DARPA, and Dr. John McTague, Deputy Director of the Office of Science and Technology Policy (OSTP) were distinguished guests at this meeting.

A list of scientific and technical problems was presented by the mentors. The topics of principal interest to the participants were (1) technologies for surveillance and verification of arms control agreements; (2) sub-orbital planes; (3) sensing (including radar and infrared) and image processing; (4) technology of anti-terrorism. Of secondary interest were the topics of (5) anti-satellite weapons; (6) strategic and battlefield C³I; (7) cryptography; and (8) stealth technology.

2. Overview Session

The purpose of the second meeting was to present a technical and programmatic introduction to some of the topics selected by the participants at the December meeting (namely, Command, Control, Communications and Intelligence (C³I); Technologies for

Topical Areas Covered (1/2- to 2-day programs)	
<ul style="list-style-type: none"> • Strategic and Tactical C³I • Surveillance and Verification of Arms Control Agreements 	<ul style="list-style-type: none"> • Strategic Computing • Sensors and Image Processing • Aerospace Plane
Visits to defense Facilities	
<ul style="list-style-type: none"> • Army Night Vision and Electro-Optics Center <ul style="list-style-type: none"> - Tactical Sensors and Supporting Technologies • National Military Command Center <ul style="list-style-type: none"> - Strategic C³ • NORAD Headquarters--Cheyenne Mountain <ul style="list-style-type: none"> - Early Warning 	<ul style="list-style-type: none"> • Falcon Air Force Station <ul style="list-style-type: none"> - National Test Bed • Buckley Air National Guard Base <ul style="list-style-type: none"> - Satellite Operations • Martin-Marietta <ul style="list-style-type: none"> - Robotics
Independent Study	
<ul style="list-style-type: none"> • Composite Materials • Mercury-Cadmium-Telluride Detectors • Missile Volume as a Possible Arms Control Constraint 	<ul style="list-style-type: none"> • Spoofing of OTH Radars • Imaging

Figure 3. Initial Year of Activities

Surveillance and Arms Control Verification; the Aerospace Plane; and Sensors and Image Processing). In addition, at the request of DARPA, a presentation was included on the DARPA Strategic Computing program.

At the conclusion of this session three topics were chosen to be explored in greater detail at the first working session. They were (1) monitoring of Soviet missile tests; (2) survivability of integrated C³I; and (3) evolution of space launch capabilities: implications for the aerospace plane.

3. First Working Session

The first summer session was divided into two parts. The first part was to present more detail to the participants about the three topical areas of interest identified at the end of the April meeting. The second part was to begin on-site visits to military and other defense facilities.

The group first visited the U.S. Army Night Vision and Electro-Optics Center at Ft. Belvoir to see a demonstration and hear about the development of tactical sensors. They also visited the National Military Command Center at the Pentagon. Expanding on the theme of the command, control and communications of our strategic nuclear forces, the group began an intensive four-day series of on-site visits to Air Force facilities. Briefings and tours were presented at the Strategic Air Command (SAC) Headquarters at Offutt Air Force Base in Omaha; and at North American Aerospace Defense Command (NORAD) Operations Center at the Cheyenne Mountain Complex, Peterson Air Force Base, Falcon Air Force Station, and Buckley Air National Guard Base, all in Colorado. The visits ended at Martin-Marietta outside of Denver with a tour of their robotics research facilities, autonomous land vehicle, and an experimental moving test bed for the development of artificial intelligence systems and advanced computer architectures.

4. Second Working Session

The purpose of this working session was to provide time to identify and study outstanding technical problems suggested by the previous sessions. Most of the effort during this session went into exploring issues dealing with imaging concepts for military sensors, missile volume as an arms control constraint, composite materials, mercury cadmium telluride detectors and radar countermeasures.

DSSG members worked individually or in small groups. This led to a flexible organization and multidisciplinary approach to problems considered. DSSG members often worked on more than one topic. While some independent study was pursued, other activities focused on developing proposals for studies and analyses in subsequent years.

Dr. Herbert York addressed the members as a special invited speaker. He discussed his recent conclusions on U.S.-Soviet offensive and defensive initiatives over the past 40 years.

5. Review Session

This session was directed at preparing and presenting the results of all technical activities for the year to IDA, DARPA and the program's mentors. Discussions were held on the strengths and weaknesses of the program and on plans for FY 1987.

B. SECOND YEAR

During the second year, the DSSG met four times, as shown in Figure 4. The activities pursued during this year are summarized in Figure 5. Five participants were added to the group, increasing the number of active participants to 17. Two new mentors were added to replace two that left. Detailed information on the DSSG activities for FY 1987 can be found in IDA Memorandum Report M-467, May 1988, and are summarized by session as follows.

Spring Technical Meeting	May 8-9, 1987
Summer Tour of Naval Operations	June 21-July 2, 1987
Fall Working Session	September 17-22, 1987
Winter Review Session	December 4-5, 1987

Figure 4. Schedule for 1987

1. Spring Technical Meeting

The purposes of the Spring meeting were to introduce the DSSG to:

(1) U.S. naval operations, technology and capabilities. Briefings on the U.S. Navy were designed to provide a technical framework in preparation for a tour through naval facilities, scheduled for the next session. Presentations were given on such inter-related topics as an overview of the U.S. Navy, technology in naval operations, the Soviet naval threat, U.S. maritime strategy, naval submarine operations, and research and development in the U.S. Navy.

(2) The role of the legislature, Congress, the Office of Management and Budget and the White House Office of Science and Technology Policy in national defense technology issues through their impact on defense funding and policy.

Topical Areas Covered (1/2-day programs)	
<ul style="list-style-type: none"> • Naval Operations and Capability • Executive Office and Congressional Views on Defense Budget and Policy • Surface Ship Torpedo Defense 	<ul style="list-style-type: none"> • Space Debris • Submarine Laser Communications • Anti-Terrorism • Superconductors
Visits to Defense Facilities	
<ul style="list-style-type: none"> • National Ocean Systems Command (NOSC) R&D on submarine laser communications, bioluminescence, microelectronics, ice phenomena, and marine mammals • Operations aboard frigates under way • Tours of an aircraft carrier and a guided missile destroyer • Miramar Navy Fighter Weapons School 	<ul style="list-style-type: none"> • Special Warfare Amphibious Base (SEALS) Trident Submarine Base • CINCPAC and CINCPACFLT • Antisubmarine warfare squadron • Naval intelligence facilities
Independent Study	
<ul style="list-style-type: none"> • Monitoring Underground Tests • Mechanical Applications of HT Superconductors 	<ul style="list-style-type: none"> • IR Detector Technology • Blue Green Laser Communications

Figure 5. Second Year of Activities

2. Summer Tour of Naval Operations

The DSSG conducted a 12-day tour to the various facilities of the U.S. Pacific Command in California, Washington State and Hawaii. The site visits and briefings were designed so that the DSSG could receive a comprehensive overview of many of the current capabilities and deficiencies of the U.S. Navy.

Visits and briefings in California included the Naval Ocean Systems Center (NOSC), the NAS Miramar flight training school, a frigate at sea, an aircraft carrier, a guided missile destroyer, an antisubmarine warfare aircraft, and the special warfare amphibious base.

At the Naval Submarine Base in Bangor, Washington, the group was provided a tour of a Trident submarine, the Trident Refit Facility, Trident Training Facility, and the Strategic Weapons Facility.

In Hawaii the group visited the Pacific Commands of CINCPAC and CINCPACFLT. Submarine Base Pacific provided a tour of an attack submarine and its training facility. In addition, visits were made to NOSC Hawaii, TACAMO at NAS Barbers Point, and two intelligence facilities.

3. Fall Working Session

The purpose of this working session was to focus the DSSG on topics for study and analysis that might be pursued in the third year of the program. Five topics were selected by the DSSG members for closer examination: surface ship torpedo defense, space debris, submarine laser communications, anti-terrorism and superconductors. These topics emanated from technical problems suggested from previous sessions, site visits or areas in which DARPA expressed an interest.

4. Winter Review Session

During this session, participants summarized the activities during the year, related reactions to what was seen and learned, presented results of research into topics in various phases of exploration, and discussed future plans and directions with IDA program staff, the mentors and DARPA representatives.

C. FINAL YEAR

During the third year, the DSSG met three times for formal sessions, as shown in Figure 6, and several participants also took part in an optional visit in the early part of the year. The activities pursued during this year are summarized in Figure 7.

NASA Johnson Space Center (Optional Visit)	March 17-18, 1988
First Summer Working Session	June 12-23, 1988
Second Summer Working Session	August/September 1988
Final Review Meeting	November 1988

Figure 6. Schedule for 1988

<p>Visits to Defense Facilities</p> <ul style="list-style-type: none"> • Army National Training Center • Nevada Test Site
<p>Independent Study</p> <ul style="list-style-type: none"> • Spoofing and Countermeasures • Opportunities for Far IR Detectors • Monitoring Underground Tests • Artillery Launched, Battlefield Surveillance System • Some Promising Directions in Applied Mathematics

Figure 7. Third Year of Activities

The principal objectives in the third year of the DSSG were to (1) provide the participants with adequate time to conduct and complete defense-related studies and analyses; and (2) help further the participants' continued involvement in defense activities upon completion of the program.

1. Optional Visit to NASA

A special opportunity was presented for interested participants to visit the NASA Johnson Space Center in Houston, Texas in March. Six participants received a tour through the Mission Control Center (MCC) and the Shuttle Avionics Integration Laboratory for testing of new hardware and software before being put into the shuttle. In addition, the DSSG received a real-time data system demonstration, and briefings on MCC data systems management, shuttle mission operations, reconfiguration of shuttle software between flights, shuttle fault isolation and malfunction procedures, and redundancy management of shuttle systems.

2. First Summer Working Session

The DSSG spent two days at the U.S. Army National Training Center (NTC) at Ft. Irwin, CA, to view unrehearsed training warfare of U.S. infantry and armor battalions against the NTC's Opposing Force (OPFOR) battalion. The OPFOR are trained in Soviet tactics and use an array of Soviet and Soviet-modified equipment in realistic scenarios

against U.S. strategy and equipment. Results of engagements are collected by video cameras and computers to assist units in improving battle readiness.

An intensive one-week summer study immediately followed the visit to the NTC. The summer study was held at the Lawrence Livermore National Laboratory (LLNL) in California. Participants decided on topics for analysis either singly or in groups and pursued their research. In addition, interested participants received tours of LLNL's facilities for microfabrication, isotope separation, the free electron laser and the NOVA laser. A special two-day tour of the Nevada Test Site was added to the agenda at the close of the summer study. Attending participants received briefings and visited the various test and fabrication facilities, including the horizontal test tunnel.

3. Second Summer Working Session

DSSG participants returned to IDA for nine days between August and September to complete analyses started during the summer study at LLNL.

4. Final Review Meeting

A final review session for the participants was held in November. In attendance were the DSSG mentors, DARPA representatives and IDA program staff. A survey of activities pursued during the two previous sessions was presented with emphasis on work in radar applications and countermeasures, monitoring underground tests, materials for far infrared detectors, and battlefield surveillance. Other efforts over the three years were also reviewed. In addition, the participants presented a program review including both impressions and a critique. The mentors, in turn, presented their impressions of the program. A roundtable discussion was held with participants, mentors and IDA program staff, concluding the session.

III. STUDIES AND ANALYSES

A significant portion of the program was devoted to the independent investigation of a variety of issues and problems that were uncovered over the three year period. The primary purposes for these investigations were to give participants the opportunity to learn to work with the defense community, apply technical skills outside of individual specialties to national defense problems, establish contacts in the defense community for the future, and work with each other in multidisciplinary groups. A secondary purpose was to give participants the chance to contribute meaningfully to the solution of outstanding defense problems or pinpoint promising areas of relevant research where possible.

The various purposes, most of which were educational in nature, coupled with the very limited amount of time allocated to studies and analyses (about a half of the approximately 20 days a year that participants devoted to the program), severely limited the comprehensiveness and rigor the participants normally expect of their own work. In some cases, such as sections on underground nuclear test monitoring and IR detectors, the reports include substantial work beyond the 20 days per year allocated for the program. In other cases, such as sections on charged particle beams and applied mathematics, the analyses were responsive to specific requests from DARPA. In all cases, the reports of efforts undertaken by the DSSG are as substantive and complete as time permitted, and are intended to be guides to promising research areas rather than completed tasks.

The sections that follow represent the results of the studies and analyses that were reported by DSSG participants at the end of 1988.

**A. MONITORING YIELDS OF UNDERGROUND NUCLEAR TESTS
USING HYDRODYNAMIC METHODS**

F. K. Lamb

Program in Arms Control, Disarmament, and International Security
and Departments of Physics and Astronomy
University of Illinois at Urbana-Champaign
and
Center for International Security and Arms Control
Stanford University

to appear in

Nuclear Arms Technologies in the 1990s
Edited by D. Schroerer and D. Hafemeister
AIP Conference Proceedings
(New York: American Institute of Physics, 1988)

PREFACE

The report that follows describes some of the results that have emerged from a study of technologies for on-site verification of nuclear test bans. This study grew out of my participation in the Defense Science Study Group organized by IDA. I first became interested in the subject as a result of a briefing on test ban verification presented by Professor Paul Richards of Columbia University at the initial Spring Meeting of the DSSG, which took place in April 1986, and by subsequent conversations with Professor Richards both at the Spring Meeting and later that year. During these conversations I became aware that DARPA is intensely interested in this subject. As a result, I chose to spend part of the DSSG summer study session held in September 1986 becoming more familiar with the issue. I also arranged for Professor Richards to visit my home institution, the University of Illinois, for two days in October 1986, during which we discussed the subject further. I then devoted part of my research effort during the Fall and Winter to increasing my understanding how the yields of underground nuclear explosions can be estimated from the shock waves they produce.

In February 1987, I was asked to brief the members and staff of the Senate Foreign Relations Committee on on-site verification methods based on shock-wave measurements. Since then I have pursued this question both at the University of Illinois and during DSSG study sessions. My involvement has led to a report that was included in the record of the Senate Foreign Relations Committee hearings on the Threshold Test Ban Treaty and Peaceful Nuclear Explosions Treaty, to service as a consultant for the Office of Technology Assessment, to the writing of a chapter for the OTA report *Seismic Verification of Nuclear Testing Treaties* (May 1988), to technical working papers issued by the University of Illinois Program in Arms Control, Disarmament, and International Security, and to a variety of other activities. The report reprinted here appeared as a chapter in *Nuclear Arms Technologies in the 1990s* (New York: American Institute of Physics, 1988).

My study of technologies for verifying limits on underground nuclear testing is continuing as an IDA task and as research at the University of Illinois supported in part by a DARPA contract through the Air Force Geophysics Research Laboratory.

MONITORING YIELDS OF UNDERGROUND NUCLEAR TESTS USING HYDRODYNAMIC METHODS

F. K. Lamb

Program in Arms Control, Disarmament, and International Security
and Departments of Physics and Astronomy
University of Illinois at Urbana-Champaign

and

Center for International Security and Arms Control
Stanford University

ABSTRACT

The yields of nuclear explosions can be estimated using hydrodynamic methods. The approach that has been proposed by the United States for nuclear test ban monitoring makes use of the fact that the initial speed of the expanding shock wave produced by an underground explosion increases with the yield. Several techniques have been developed in the United States to measure the speed of the shock wave, of which the so-called CORTEX technique is the most recent and best. A variety of algorithms have been used to derive yield estimates from shock wave radius vs. time measurements. Although more intrusive than seismic methods, current hydrodynamic methods could be used to monitor the Threshold Test Ban and Peaceful Nuclear Explosion treaties, provided that appropriate changes in these treaties are negotiated and that adequate cooperative arrangements are made to assure accuracy. Significant engineering, operational, and analysis problems need to be solved before these methods could be used to monitor with confidence a low-threshold test ban. The methods are not relevant to a comprehensive test ban.

I. INTRODUCTION

Hydrodynamic methods have long been used to estimate the yields of nuclear explosions, both in the atmosphere¹ and underground.² All such methods are based on the fact that the strength of the shock wave produced by an explosion increases with the yield. Hydrodynamic methods were introduced as a treaty-monitoring tool in the Protocol of the Peaceful Nuclear Explosions Treaty (PNET) of 1976, which explicitly established such methods as among those that could be used to monitor the yield of any salvo of explosions with a planned aggregate yield greater than 150 kilotons (kt).³ Hydrodynamic methods have recently become a focus of attention as a result of controversy over monitoring of the Threshold Test Ban Treaty (TTBT).

The TTBT was signed by the United States and the Soviet Union in 1974 and banned underground nuclear tests with yields greater than 150 kt after March 31, 1976.³ Although neither party has ratified the TTBT, both have separately stated that they will respect the 150 kt limit. From that time to the present, the U.S. government has relied primarily on yield estimates based on teleseismic body wave magnitudes measured outside the Soviet Union to monitor the TTBT limit.⁴ According to the U.S. Department of State, this method has an uncertainty of approximately a factor of two at the 95% confidence level for Soviet tests with yields near 150 kt.^{5,6} This statement means that 95 times out of 100 the estimated yield of a 150 kt explosion will lie between 75 kt and 300 kt. Put another way, there is only one chance in 40 that a single explosion with a yield of 300 kt would appear to have a most likely yield of 150 kt or less. The probability that two such explosions would appear to be treaty compliant is 1 in 1600, and so on.⁷ According to recent studies, the best seismic methods now available are expected to have an uncertainty of a factor of 1.5-1.6 at the 95% confidence level for explosions with yields above 50 kt, if measurements are made only outside the Soviet Union and the Soviet test site is not calibrated, or a factor of 1.3, once the Soviet test site is properly defined and calibrated.⁴

Prior to the Reagan administration, the precision of remote seismic methods was considered adequate. However, the Reagan administration has stated that "the remote seismic techniques we must rely on today to monitor Soviet nuclear tests do not provide yield estimates with the accuracy required for effective verification of compliance."⁵ As an alternative to these techniques, the Reagan administration has since 1983 strongly advocated routine use of hydrodynamic methods to monitor the TTBT.^{5,6} It believes that it has identified in hydrodynamic yield estimation methods an approach "which will reduce the uncertainty in yield measurement to an acceptable level and will do so without danger of compromising other sensitive information about the nature or performance of the nuclear device whose yield is to be measured."⁶

The present article reviews hydrodynamic yield estimation methods and their application to test ban monitoring. §II provides a brief overview of the development of an underground nuclear explosion and the variety of yield estimation methods currently in use. §III summarizes some relevant material properties of rock and describes the evolution of the shock wave produced by an underground nuclear explosion. In §IV we explain the technique advocated by the Reagan administration to measure the evolution of the shock wave. In §V we describe the algorithms that are currently used to derive yield estimates from these measurements. Technical issues related to the use of hydrodynamic methods to monitor test ban agreements are considered in §VI. In §VII we briefly compare hydrodynamic and seismic yield estimation methods, discuss the current status of U.S.-Soviet test ban negotiations, and mention several public policy issues raised by the U.S. drive to gain Soviet acceptance of hydrodynamic methods for routine monitoring of the TTBT. Our conclusions are summarized in §VIII.

II. MONITORING UNDERGROUND NUCLEAR EXPLOSIONS

Currently, about 90% of U.S. nuclear tests are conducted in vertical shafts at depths of 250–650 m; the remainder are conducted in tunnels.⁸ In preparing for a test that will be conducted deep underground, the vertical hole that will contain the nuclear charge (the so-called emplacement hole) is first drilled, a process that typically takes 8–10 weeks. The charge and diagnostic equipment are then placed in canisters, which can be as much as 15 m in length, and lowered into the emplacement hole. Depending on the nature of the test, a variety of diagnostic pipes and cables may lead upward from the canisters to the surface (current U.S. nuclear weapon tests involve anywhere from ~50 to ~250 diagnostic pipes and cables). After the canisters are in place, the hole is stemmed with sand, gravel, and plugs.⁹

In the following two subsections we first describe what happens when a nuclear charge is exploded deep underground and then explain briefly the methods that are used to estimate the yields of such explosions.

A. Phases of an Underground Explosion

For present purposes the time development of the explosion may be divided into the following three somewhat simplified phases:^{10,11}

Initial phase.—The release of nuclear energy is accompanied by emission of nuclear radiation, fission fragments, and thermal electromagnetic radiation. The temperature in the nuclear charge rises steeply, reaching 10^7 K within a microsecond or so. At the very earliest times, the energy of the explosion is carried outward by the expanding weapon debris and radiation. Soon, however, the vaporized nuclear charge and nearby rock form a bubble of hot gas in which the initial pressure is hundreds of Mbar. The enormous pressure in the bubble causes it to expand rapidly, creating a cavity and driving a shock wave into the surrounding rock. The final radius R_c of the cavity depends somewhat on the depth of the explosion and the composition of the surrounding rock, as well as the yield. For a burst of yield W , a useful approximate expression is¹²

$$R_c \approx 14 (W/1 \text{ kt})^{1/3} \text{ m.} \quad (1)$$

The cavity reaches its final radius in about $90 (W/1 \text{ kt})^{1/3}$ ms.

Hydrodynamic phase.—The shock wave initiated by the expansion of the hot gas propagates outward at a speed that is initially much greater than the speed of sound in the surrounding, undisturbed rock. At this early time, the stress produced by the shock wave greatly exceeds the critical stress at which the rock becomes plastic, so that to a good approximation the rock can be treated as a fluid. This phase is therefore referred to as the “hydrodynamic” phase. (In defining the hydrodynamic phase, we emphasize the prefix *hydro* and simply require that the shocked rock behave like a fluid. Other authors emphasize instead the root *dynamic* and require not only that the shocked rock behave like a fluid, but also that the speed of the shock wave greatly exceed the speed of

sound in the rock; this second usage is common in the Soviet Union.) As the shock wave continues to expand, it weakens. Eventually, the strength of the rock can no longer be neglected. This marks the end of the hydrodynamic phase.

Final phase.—Even after the compression wave is no longer hydrodynamic, the rarification wave that follows is still strong enough to fracture rock. Intense fracturing typically occurs out to a radius $\sim 3R_c$.¹² Beyond this point, the degree of fracturing caused by the expanding shock wave drops dramatically until, at $\sim 5R_c$, fracturing essentially stops. (Rarification waves caused by reflection of the shock wave from the surface or collapse of the roof of the cavity may cause fracturing beyond this radius.) The shock wave then continues to expand nearly elastically, eventually evolving into the leading wave of a train of elastic (seismic) waves. These waves, which typically carry $\lesssim 5\%$ of the energy of the explosion, propagate through and around the earth and can be observed at points thousands of kilometers from the site of the explosion.

B. Monitoring Methods

Three different types of methods are commonly used to estimate the yields of nuclear weapon tests. These types make use of phenomena that occur during the three phases of the explosion identified above.

Radiochemical methods make use of the nuclear reactions that occur during the first phase of the explosion. By knowing the relative abundances of various nuclides in the original nuclear device and by determining the relative abundances of fission fragments and fusion products after the explosion, the yield of the explosion can be estimated. However, at present there are several barriers to using radiochemical methods to monitor test ban treaties. First, the monitoring party must be able to make a variety of measurements at the test site. Second, to achieve high precision with this method some knowledge of the design of the nuclear charge may be required. Third, in addition to the yield, radiochemical methods can provide other information about the design and performance of nuclear devices, which may be considered sensitive. For these reasons, radiochemical methods are not usually considered for treaty monitoring.¹³

Hydrodynamic methods make use of the fact that the strength of the shock wave produced by an explosion increases with the yield, other things being equal. As a result, the peak particle velocity, pressure, and density are greater at a given radius for explosions of greater yield. By comparing measurements of these quantities with a model of the evolution of the shock wave based on knowledge of the nature and structure of the geologic media in which the explosion occurred, the yield of the explosion can be estimated. In order to use hydrodynamic methods to monitor test ban treaties, the monitoring party must have access to the test site in order to determine the relevant properties of the geologic media there before the test and to measure the evolution of the shock wave during the test. To assure high accuracy, constraints on the test geometry are also required. Hydrodynamic methods are not part of current TTBT verification provisions.

Seismic methods make use of the ground motions caused by the elastic waves that propagate through and around the earth during the third stage of an underground nuclear explosion. Some seismic waves (such as body and surface waves) propagate to so-called teleseismic distances ($\geq 2,000$ km) from the explosion. Yield estimation methods based on these waves can be used with measurements of ground motion made at stations outside the country in which the test occurs. Other seismic waves (such as L_g waves) typically propagate only to regional distances ($< 2,000$ km). Yield estimation methods based on these latter waves may require data from in-country stations. To assure high accuracy, knowledge of the geologic media at the test site and the way in which the earth near the site transmits seismic waves is required. As noted earlier, the United States routinely uses seismic data taken at stations outside the Soviet Union to monitor the TTBT. In-country monitoring stations and independent access to data on the seismic properties of the test site are not part of the current verification provisions of the TTBT. For recent reviews of seismic methods, see refs. 4 and 14-18.

It has been claimed several times in recent Congressional hearings on TTBT and PNET verification¹⁹⁻²³ that hydrodynamic methods are "direct" whereas seismic methods are not. From a scientific point of view there is no such distinction. All three methods just described involve (1) production of a signal by the exploding charge, (2) propagation of the signal to locations more or less remote from the detonation point, and (3) detection of the signal by sensors at those locations. Important questions are how the size of the signal varies with yield, how well the propagation of the signal is understood, and how accurately and precisely the sensors can measure the signal.

It has also been asserted²² that use of hydrodynamic methods in and of itself eliminates the possibility of systematic error or "bias". Obviously it does not. All three methods are subject to both systematic and random errors. Relevant questions are the expected sizes of the errors, and whether they are so large as to be of concern.

III. SHOCK WAVE EVOLUTION

In hydrodynamic methods of yield estimation, the size of the explosion is estimated by fitting a model of the evolution of the shock wave, which depends parametrically on the yield, to measurements of the motion. Shock waves in rock behave differently from shock waves in air primarily because the atoms in rock are close together and interact strongly.²⁴ Therefore, in discussing the application of hydrodynamic methods to underground explosions it will be helpful to have in mind some relevant material properties of rock as well how shock waves produced by underground explosions evolve in rock.

A. Rock Properties

The strength of a shock wave can be characterized by the peak pressure that it produces. Weak shock waves and acoustic waves in rock propagate at a

constant speed, the so-called elastic wave speed²⁵

$$c_t = \left(\frac{K_0 + \frac{4}{3}G_0}{\rho_0} \right)^{1/2} \quad (2)$$

Here K_0 and G_0 are the bulk and shear moduli, respectively, of the rock in its standard state, and ρ_0 is the mass density. For granite, $K_0 \approx 360$ kbar and $G_0 \approx 320$ kbar,²⁶ giving $c_t \approx 5 \text{ km s}^{-1}$.

Shock waves that are strong enough to produce stresses in excess of the critical shear stress p_{crit} cause the rock to lose its firmness and to become plastic (for granite, p_{crit} is about 40 kbar for high strain rates²⁶). Such waves are called plastic waves. The speed of a plastic wave increases with its strength. The weakest such waves propagate at the low-pressure plastic wave speed²⁵

$$c_0 = \left(\frac{K_0}{\rho_0} \right)^{1/2}, \quad (3)$$

which is determined by the compressibility of the rock in its standard state. Since only the bulk modulus contributes to c_0 , it is necessarily less than c_t . For solid granite, $c_0 \approx 4 \text{ km s}^{-1}$.

In the hydrodynamic regime, conservation of mass, momentum, and energy across the shock front imply²⁷

$$\epsilon(p, V) - \epsilon_0(p_0, V_0) = \frac{1}{2}(p_0 + p)(V_0 - V), \quad (4)$$

where ϵ_0 and ϵ , p_0 and p , and V_0 and V are, respectively, the internal energies, pressures, and specific volumes ahead of and just behind the shock front. By analogy with the equation relating the initial and final pressures and volumes during adiabatic compression of a fluid, this relation, which is of the form

$$p = H(V, p_0, V_0), \quad (5)$$

is called the *shock adiabat* or *Hugoniot*.

Conservation of momentum across the front of a hydrodynamic shock wave implies

$$p = p_0 + \rho_0 D u, \quad (6)$$

where D is the speed of the shock front, measured in the rest frame of the undisturbed rock, u is the particle speed just behind the shock front, and we have assumed that the rock in front of the shock front is at rest. Thus, the Hugoniot may be expressed as a relation between D and u , that is

$$D = H(u). \quad (7)$$

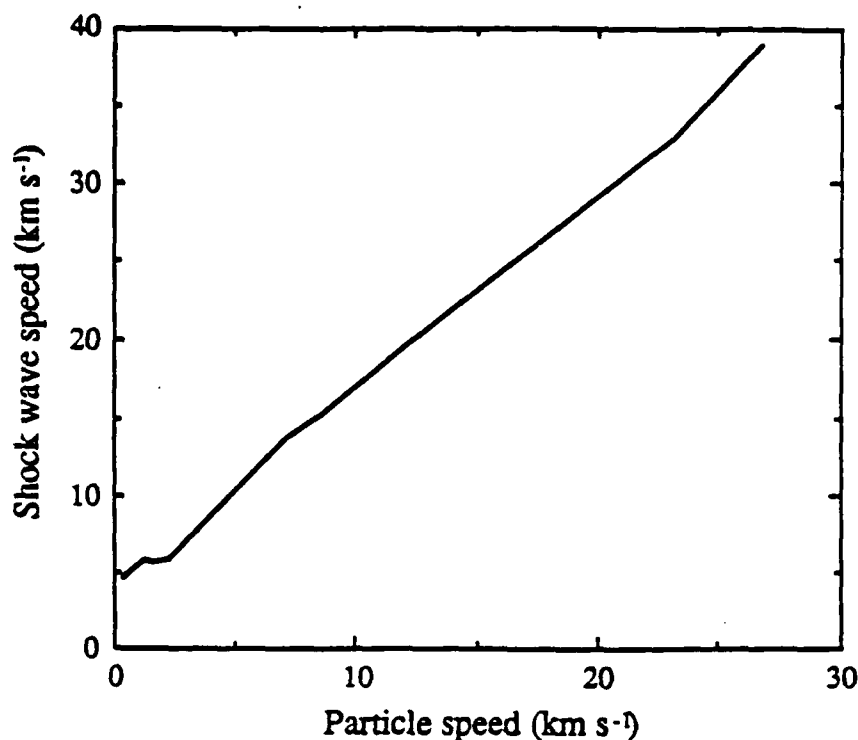


Fig. 1.—Relation between shock wave speed D and particle speed u just behind the shock front for granite (ref. 29). Note the approximate linearity of the Hugoniot at large u . The step in the curve at $u \approx 2 \text{ km s}^{-1}$ reflects a phase transition that occurs at about 400 kbar.

Figure 1 shows a recent Hugoniot for granite expressed in this way. In general, the Hugoniot depends on the chemical composition, porosity, gas-filled porosity, fracture pattern, degree of liquid saturation, and other properties of the rock.

For some rocks, the Hugoniot may be adequately represented in the *hydrodynamic regime* by a single linear relation of the form^{27,28}

$$D = A + Bu, \quad (8)$$

which implies the Hugoniot curve²⁷

$$p = \frac{A^2(V_0 - V)}{(B - 1)^2 V^2 \left[\frac{B}{B-1} - \frac{V_0}{V} \right]^2}. \quad (9)$$

Table 1 lists values of A , B , and ρ_0 for granite and wet tuff that were derived by fitting a Hugoniot of the form (9) to recent high-pressure equations of state for these materials.

For Hugoniots of the form (9), the ratio ρ/ρ_0 of the material density immediately behind the shock front to the material density ahead of the shock front increases with the strength of the shock wave until it reaches a certain value $(\rho/\rho_0)_{\max} = (V_0/V)_{\max} = B/(B - 1)$. Once the wave has become this strong,

TABLE 1
Rock Equations of State^a

Rock	ρ_0 (g cm ⁻³)	A (km s ⁻¹)	B	$R_{t,1}$ (m)	$R_{t,150}$ (m)
Granite	2.67	2.80	1.45	3.8	20
Wet tuff	1.95	1.45	1.62	7.0	37

^aThe parameters ρ_0 , A , and B are from ref. 30 and were obtained by fitting a Mie-Grüneisen equation of state to recent tabulated equations of state for granite (ref. 29) and wet tuff (ref. 31) at high pressures. $R_{t,1}$ and $R_{t,150}$ are characteristic shock wave transition radii (see text) for 1 kt and 150 kt explosions.

any further increase in its strength does not produce any increase in the ratio ρ/ρ_0 . For this reason, the density ratio $(\rho/\rho_0)_{\max}$ is referred to as the limiting density ratio. For the granite Hugoniot listed in Table 1, the limiting density ratio is ~ 3 . Peak pressures ~ 10 – 100 Mbar are required to achieve density ratios near the limiting value.

If a single linear relation adequately describes the Hugoniot at large u and if this relation could be extrapolated to small u , the constant A would correspond to the low-pressure plastic wave speed c_0 . However, the large- u relation usually is not valid for small u , and hence A usually does not equal c_0 . In granite, for example, A is about 3 km s^{-1} whereas c_0 is about 4 km s^{-1} .

Even if the Hugoniot is not linear over the range of u that is of interest, a curve consisting of piece-wise linear segments of the form (8) may serve as a practical approximation to $H(u)$ for many purposes.²³

B. Simplified Model

To understand the evolution of shock waves produced by underground nuclear tests, it is helpful to consider first the shock wave that would be produced by release of a large amount of energy in an infinitesimal spherical volume (a so-called point explosion) in homogeneous rock. A qualitative understanding of the development of the shock wave produced by such an explosion can be gained from the following relatively simple model. Suppose we make the *ansatz* that throughout the motion the particle speed u just behind the shock front is related to the yield W of the explosion by the expression^{30,32}

$$u(t) = \left(\frac{3fW}{4\pi R^3 \rho_0} \right)^{1/2}, \quad (10)$$

where R is the radius of the shock front and f is a constant dimensionless factor that describes how the energy of the explosion is partitioned between kinetic energy of bulk motion and internal energy, and how the velocity, density, and internal energy of the shocked material vary with position. The factor f is constant for a self-similar shock wave (see below) but is not expected to remain constant

as the shock wave evolves.³² Nevertheless, shock wave radius and particle speed data from actual underground nuclear tests as well as computer simulations of such tests indicate that relation (10) with $f \approx 0.5$ is satisfied fairly well for granite and wet tuff until relatively late times.^{30,32}

Let us now assume, for the sake of illustration, that the Hugoniot of the medium can be adequately represented by a *single* linear relation of the form (8) over the whole range of u that is of interest. Then $c_0 = A$. Therefore, in the following discussion we refer to A as the low-pressure plastic wave speed.

Given the *ansatz* (10), the Hugoniot can be rewritten as³²

$$D \equiv \frac{dR}{dt} = A + Bu = A \left[\left(\frac{R_t}{R} \right)^{3/2} + 1 \right], \quad (11)$$

where

$$R_t \equiv \left(\frac{3fWB^2}{4\pi\rho_0 A^2} \right)^{1/3} \quad (12)$$

is the characteristic transition radius that separates the region where $D \propto R^{-3/2}$ from the region where $D \approx A$. Typical values of R_t for 1 kt and 150 kt explosions in granite and wet tuff are listed in Table 1 for the values of A and B given there. In dry alluvium, R_t can sometimes exceed 60 m for a 150 kt explosion.

Equation (11) can be integrated to obtain a simple, closed expression for $R(t)$, from which one can calculate $D(t)$, $u(t)$, $p(t)$, and $\rho(t)$.^{32,33} This model shows in a qualitative way how the evolution of the shock wave depends on the yield of the explosion and the equation of state of the rock. As an example, the peak pressure, peak density, and radius of the shock front at various times are listed in Table 2, for 1 kt and 150 kt explosions in granite. The pressure p_0 of the overburden is ~ 200 bar for a depth of 1 km and hence can be neglected for all depths and times of interest.

The motion of the shock front given by equation (11) can be divided into three different intervals:

Strong Shock Interval.—Initially, the speed of the shock front is much greater than the speed of sound in the undisturbed rock, the pressure behind the shock front is predominantly thermal pressure, and the ratio of the density immediately behind the shock front to the density ahead of the front is close to its limiting value. This is the strong shock interval.²⁴

For a spherically-symmetric point explosion in a homogeneous medium, the motion of the shock wave in the strong shock interval is self-similar.³⁴⁻³⁶ In such a motion, the distributions with radius of the pressure, density, and particle velocity evolve with time in such a way that only their scales and the radius of the shock front change, while the shapes of the distributions remain unaltered. The evolution of such a shock wave is only weakly dependent on the properties of the medium. Thus, simple models can be used to estimate the yield if there is an interval of self-similar motion and if data from this interval are available.

TABLE 2
Shock Wave Evolution in Granite^a

Pressure (kbar)	Density (ρ_{\max})	1 kt Explosion		150 kt Explosion	
		Time (μ s)	Radius (m)	Time (μ s)	Radius (m)
70,000	0.9	4	0.5	20	3
10,000	0.8	10	0.9	80	5
4,000	0.7	40	1.4	200	8
1,500	0.6	90	2	500	11
500	0.5	200	3	1,200	17
150	0.4	600	5	3,000	30

^aFor the model of a spherically-symmetric point explosion described in the text. The Hugoniot (8) was used, with the values of A and B given in Table 1. The phase transition that occurs at ~ 400 kbar (cf. Fig. 1) has been neglected. The unit of density, ρ_{\max} , is the limiting density of granite (see text), which is 9.4 g cm^{-3} for this equation of state.

For example, in the evolution given by equation (11), the motion of the shock front radius R during the strong shock interval ($R \ll R_t$) satisfies

$$R(t) \approx W^{1/3} \left(\frac{75fB^2}{16\pi\rho_0} \right)^{1/5} \left(\frac{t}{W^{1/3}} \right)^{2/5} \propto W^{1/3} \left(\frac{t}{W^{1/3}} \right)^{2/5}. \quad (13)$$

Thus, the exponent of time is independent of the properties of the medium. Moreover, the evolution is only affected by ρ_0 and B .

Unfortunately, in actual underground nuclear tests self-similar motion does not have time to develop, given current testing practices and the yields of interest, as explained below.

Transition Interval.—As the shock wave expands, it weakens and slows, and the peak pressure and density drop. When the shock front reaches a certain radius R_1 , the peak density ratio is 0.8 times the limiting value and we say that the shock wave has entered the transition interval. (The motion of the shock wave changes only gradually and so the point at which it is said to enter the transition interval is purely conventional. Throughout the present article we use the convention that the transition interval begins when the peak density ratio falls to 80% of its limiting value.) For an explosion in granite, this occurs when the peak pressure has fallen to $\sim 10,000$ kbar (cf. Table 2). For a 1 kt explosion in granite R_1 is ~ 1 m, whereas for a 150 kt explosion R_1 is ~ 5 m.

Over most of the transition interval, the thermal pressure just behind the shock front is not much greater than the cold pressure of the compressed rock, although the speed D of the shock front is still much larger than the low-pressure

plastic wave speed A . In this interval, the motion of the shock wave is more sensitive to the properties of the medium than it is in the strong shock interval, depending on A as well as B and ρ_0 even in the simple model of equation (11). Consequently, more knowledge of the ambient rock is required in order to make accurate yield estimates using data taken in this interval.

Plastic Wave Interval.—As the shock wave expands and weakens further, the thermal pressure behind the shock front becomes a small fraction of the total pressure and the shock speed D approaches the low-pressure plastic wave speed A . At a certain radius R_2 ($\sim R_1$), the shock speed is 1.2 times the low-pressure plastic wave speed and we say that the shock wave has entered the plastic wave interval. (Again, the motion of the shock wave changes only gradually and so the point at which it is said to enter the plastic wave interval is purely conventional. Throughout the present article we use the convention that the plastic wave interval begins when the shock speed falls to 1.2 times the low-pressure plastic wave speed.) For an explosion in granite, this occurs when the peak pressure has fallen to ~ 150 kbar, corresponding to a peak density ratio ~ 0.4 times the maximum (cf. Table 2). For a 1 kt explosion in granite R_2 is ~ 5 m, whereas for a 150 kt explosion R_2 is ~ 30 m.

For the model given by equation (11), the plastic wave interval corresponds to $R \geq 3 R_1$. In this regime

$$R \approx \text{const.} + A t, \quad (14)$$

where the constant is determined by the motion in the strong shock and transition intervals.

The evolution of the shock wave in the plastic wave interval is sensitive to the equation of state and the constitutive relations describing the rock. For actual rocks, the evolution in this interval is complex, as described below. Hence relatively detailed modeling is required in order to obtain accurate yield estimates from data taken in this interval.

C. Other Effects

The evolution of the shock wave produced by an actual underground nuclear test is more complex than the evolution just described, for several reasons. First, the actual shock wave is not produced by a spherically-symmetric point explosion. The emplacement holes currently used in U.S. tests have radii R_e as large as 1.5 m^{10} and emplacement holes with larger radii are planned for the future. Moreover, the nuclear charge and diagnostic canisters may be many meters in length. As a result, the source of the shock wave is vapor and radiation filling a volume with a dimension of meters. Moreover, the explosion is usually not spherically symmetric, causing the expanding shock wave to be aspherical initially. The energy flows produced by explosions in tunnels may be even more complex.

The motion of the shock wave cannot become self-similar until the shock wave has enveloped a mass of material much greater than the mass of the nuclear charge and casing, and energy transport by radiation is negligible. The

radius R_0 at which this occurs is necessarily much greater than the radius R_e of the emplacement hole and depends on the design of the nuclear charge and surrounding diagnostic equipment. Unless there is a range of radii satisfying $R_0 \ll R \ll R_1$, the shock wave will not have time to become self-similar before entering the transition interval. The data in Table 2 show that no such range exists in granite for current emplacement practices, even for explosions as large as 150 kt. Thus, the simplicity of estimating yields from an interval of self-similar motion cannot be realized, given current testing practices and the yields of interest.

Other complications arise from the complexity of the motion in the transition and plastic wave intervals in actual rocks. As the shock wave expands and weakens, one or more phase transitions may occur. We have already noted that granite undergoes a phase transition when the peak pressure falls to ~ 400 kbar. At a somewhat lower peak pressure (~ 200 kbar in granite), the shock speed becomes less than the elastic wave speed c_t . At this point the shock wave splits into an elastic wave followed by a plastic wave.^{25,37} The initial pressure jump in the elastic wave is p_{crit} , which is ≈ 40 kbar for granite.²⁶ Since $c_0 < c_t$, the weakening plastic wave cannot overtake the elastic wave and the two-wave structure is stable. As a result, the plastic wave propagates through rock that has already been compressed and accelerated to a velocity $\sim 1-10 \text{ m s}^{-1}$ by the elastic wave. (Since c_0 is several km s^{-1} , for most purposes the acceleration of the medium by the elastic wave may be neglected and the plastic wave may be taken to propagate with a speed c_0 relative to the undisturbed ambient medium, as was done in eq. [6]). Finally, when the peak pressure in the plastic wave is no longer much greater than the critical shear stress p_{crit} , the shear strength of the rock can no longer be neglected. In granite, for example, p_{crit} is ≈ 40 kbar, and the hydrodynamic approximation therefore begins to fail when the peak pressure falls below ~ 150 kbar.

In addition to these complications that occur in a homogeneous medium, theoretical models and experimental data show that the evolution of shock waves in actual geologic media can be affected by voids, layering, and other geologic structures that may vary from one test site to another.

IV. MEASURING SHOCK WAVE EVOLUTION

A variety of sensors have been used at different times in the United States to measure the evolution of shock waves produced by underground nuclear explosions. These include strain gauges, particle velocity gauges, accelerometers, and pressure sensors.²⁶ In the United States, recent efforts to develop instrumentation for test ban monitoring have focused on radius vs. time (RVT) techniques, which measure the radius of the shock front as a function of time.

A. Using Sensing Cables

One RVT measuring technique that has been used in the United States since the early 1960s is based on the fact that the pressure peak near the front

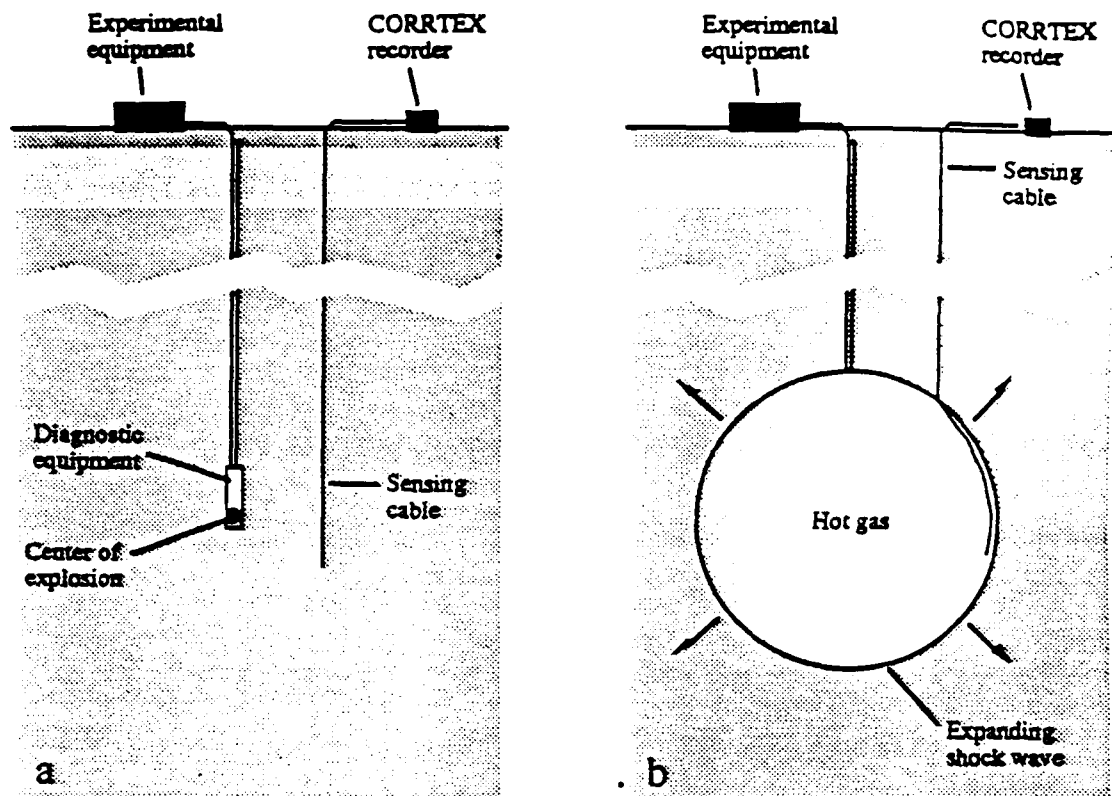


Fig. 2.—Schematic drawings illustrating (a) placement of the shock front sensing cable in a satellite hole and (b) progressive shortening of the cable by the expanding shock wave produced by a nuclear explosion.

of the shock wave can crush an electrical cable. In this approach, a coaxial sensing cable is lowered into the emplacement hole before it is backfilled or into one or more other holes (so-called satellite holes) that have been drilled nearby specifically for this purpose. The satellite-hole geometry is shown in Figure 2a. If the sensing cable is strong enough that it is not crushed by the elastic precursor or other unwanted signals but weak enough that it is crushed by the pressure peak in the hydrodynamic shock wave, the cable will be electrically shorted or its impedance substantially changed near the point where the hydrodynamic shock front intersects it. As the shock wave expands with time, the length of cable from the measuring equipment to the shallowest point at which it has been crushed is measured by electrical equipment attached to the cable and located above ground, as shown in Figure 2b. If the time at which the explosion began and the path of the cable relative to the center of the explosion are known and if the explosion is spherically symmetric, the radius of the shock front as a function of the time since the beginning of the explosion can be calculated from a record of the changing length of the cable.

As a concrete example, suppose the sensing cable is placed in a vertical satellite hole displaced laterally a distance d from the emplacement point. Then if the length L_0 of the sensing cable from the surface to the point where the cable

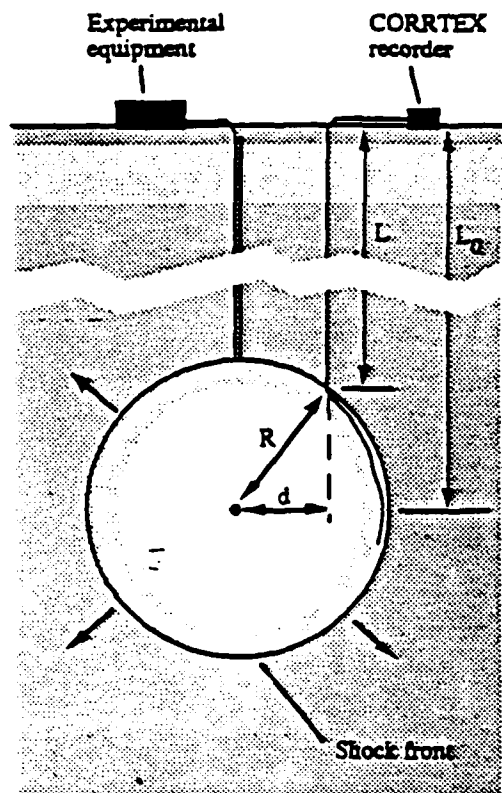


Fig. 3.—Schematic drawing illustrating how the radius R of the shock wave at time t is related to the lateral displacement d of the sensing cable from the center of the explosion, the length L_0 of the sensing cable from the surface to the point where the cable is first crushed, and the length L of the cable.

is first crushed is known, measurements of $L(t)$ can be converted into estimates of the radius R of the shock front as a function of time t using the Pythagorean relation

$$R(t) = \left[d^2 + (L_0 - L(t))^2 \right]^{1/2}. \quad (15)$$

The geometrical meaning of the quantities appearing in equation (15) is shown in Figure 3. In practice the sensing cable usually is not perfectly vertical and more complicated calculations must be performed to convert $L(t)$ to $R(t)$.

As discussed in §V, an error of 1 m in the measured distance of the crushing point from the center of the explosion will cause an error of about 50 kt in the yield estimate, for yields near 150 kt. Thus, accurate surveys of the emplacement and satellite holes and an accurate knowledge of the path of the sensing cable within the satellite hole are required in order to make an accurate yield estimate. If the cable wanders within the hole and this is not taken into account, the length of the cable crushed by the shock wave will be greater than the linear distance traveled by the shock front, causing the the speed of the shock wave and therefore the yield of the explosion to be overestimated. In the United States, hole surveys are currently made with special laser or gyroscopic equipment. In some yield

estimation algorithms, the lateral displacement of the satellite hole from the center of the explosion can be treated as one of the unknowns in estimating the yield.

If the explosion is not spherically symmetric, due to the effect of the canister, the design of the test geometry, or inhomogeneities in the ambient medium, interpretation of the sensing cable data becomes more complicated and could be ambiguous or misleading under the conditions likely to be encountered in treaty monitoring. Potential problems of this kind could be addressed by cooperative agreements, as discussed in §VI.

Sensing cables with a variety of crushing strengths, ranging from as little as 30 bar to as much as 30 kbar, have been used.³⁸ However, even the latter cable can be crushed by the elastic precursor in granite. Thus, once the shock wave has split, the length of cable to the crushing point may reflect the position of the elastic shock front rather than that of the hydrodynamic shock front which follows.^{39,40} If so, the sensing cable cannot provide data about the motion of the hydrodynamic shock wave.²⁶ Worse, if the motion is misinterpreted as that of the hydrodynamic wave, the estimated yield of the explosion will be erroneously high. As a result, interpretation of data in this interval requires great care.

The Reagan administration has strongly advocated adoption of the satellite-hole geometry for monitoring the TTBT using hydrodynamic methods. In order to collect ground shock data in the hydrodynamic region from explosions with yields near 150 kt, the satellite hole must be slightly deeper than the emplacement hole, which is typically ~650 m deep for such explosions, and the deepest portion of the satellite hole must be no further than ~30 m from the center of the explosion. In order to use the particular hydrodynamic yield estimation algorithm advocated by the administration (insensitive-interval scaling, see §V), the sensing cable must cover the algorithmic interval, which is a more demanding criterion and requires that the bottom of the satellite hole be no more than ~10 m from the center of the explosion, for yields near 150 kt.

In order to apply hydrodynamic methods to monitoring a 10 kt low-threshold test ban, the bottom of the satellite hole would have to be well within ~10 m of the emplacement point, which is typically ~250 m deep for charges with yields near 10 kt. In order to apply insensitive-interval scaling, the satellite hole would have to be within ~4 m of the center of the explosion.

Use of the satellite-hole geometry requires sophisticated drilling capabilities in order to make sure that the satellite hole maintains the proper separation from the emplacement hole.⁴¹ In addition, conversion of cable length to shock wave radius is more complicated if the cable is in a satellite hole than if it is in the emplacement hole. On the other hand, the satellite-hole geometry reduces the intrusiveness of the method and eliminates the disturbing effects of jetting within diagnostic pipes and other phenomena that can crush or short the sensing cable ahead of the hydrodynamic ground shock. In the discussion that follows, we shall assume that the sensing cable is in a satellite hole, unless otherwise stated.

B. Measuring the Position of the Crushing Point

During the 1960s and 1970s, the position of the crushing point was measured in the United States using a technique called SLIFER (which is an acronym for Shorted Location Indicator by Frequency of Electrical Resonance).^{26,42} In this approach, the cable is used as the inductive element of a resonant oscillator. As the cable is progressively crushed, the frequency of the oscillator changes. By knowing the propagation velocity of electromagnetic signals in the cable and the frequencies of the oscillator that correspond to $L = 0$ and $L = L_0$, one can convert measurements of the change in oscillator frequency during the explosion to estimates of the change in the length of the cable.

In the late 1970s, an improved approach to measuring the length of sensing cables, called CORRTEX, was developed (CORRTEX is an acronym for Continuous Reflectometry for Radius versus Time Experiments).^{39,40,43,44} In this approach, a sequence of electrical pulses is sent down the cable. At the crushing point, these pulses are reflected back up the cable to recording equipment. By knowing the speed at which the pulses propagate down and up the cable, the round-trip travel time of each pulse can be converted into an estimate of the length of the cable at the time the pulse was reflected.

Current (CORRTEX III) equipment can store up to 4,000 data points. Pulse separations from $10\ \mu\text{s}$ to $90\ \mu\text{s}$ can be selected, giving a record of the changing cable length that is 40 ms to 360 ms in length. The pulses typically propagate down and up the sensing cable at about $2 \times 10^5\ \text{km s}^{-1}$. A typical uncertainty in the round-trip travel time during a nuclear explosion is 500 ps, corresponding to an uncertainty of about 10 cm in the measured length of the cable or about 5 cm in the distance to the crushing point. The cable length measurements can be checked by creating fiducial loops in the cable at predetermined points, which will create downward jumps in the cable length as the crushing point passes over them. Using these jumps, the cable length data can be adjusted for systematic errors. The time at which the explosion begins is estimated by recording the time at which the electromagnetic pulse (EMP) from the explosion arrives at the CORRTEX recorder. The CORRTEX technique is less affected by disturbing signals produced by the explosion than were earlier techniques and is the technique that has been advocated by the Reagan administration for monitoring the TTBT.

RVT techniques are also used to estimate the yields of underground nuclear tests in the Soviet Union. The Soviets use two different sensing techniques, called Mis (or Miz) and Contactor.^{45,46} The current Soviet approach reportedly uses a sensing system with switches that are sequentially destroyed by the shock front.⁴⁷

V. YIELD ESTIMATION ALGORITHMS

Once measurements of the length of the sensing cable have been converted to estimates of the radius of the shock front as a function of time, the yield

of the explosion is estimated by applying some algorithm, by which we mean a particular procedure for comparing the RVT data with a particular model of the motion of the shock front. Because hydrodynamic methods of yield estimation are evolving as research aimed at increasing our understanding of underground explosions and improving yield estimation methods continues, the description of yield estimation algorithms in the present section is necessarily a status report.

Although the algorithms used by different individuals or groups can (and usually do) differ in detail, most of the algorithms currently in use are of four basic types: insensitive interval scaling, similar explosion scaling, semi-analytical modeling, and numerical modeling. These are the algorithms that are discussed here. In order to simplify matters, we will assume at first that the explosion is spherically symmetric and that the ambient geologic medium is homogeneous. Some of the complications that can arise if the explosion is aspherical or the medium is inhomogeneous are described at the end of this section.

A. Insensitive Interval Scaling

The simplest algorithm currently in use is insensitive interval scaling. This is the algorithm that the Reagan administration has proposed to use in analyzing CORTEX data as a new routine method of monitoring Soviet compliance with the TTBT.

Insensitive interval scaling assumes that the radius of the shock wave produced by an explosion of given yield does not depend on the rock in which the explosion occurs during a certain interval in time and radius called the "insensitive interval". This assumption is based on empirical evidence that the radius of the shock front is relatively insensitive to the medium during a certain interval in time and radius toward the end of the transition interval, for the particular geologic media for which the United States has good experimental data or theoretical models. These media include the dry alluvium, partially saturated tuff, saturated tuff, granite, basalt, and rhyolite at the nuclear test sites the U.S. has used. These media are mostly silicates and almost all are located at the Nevada Test Site. The radius of the shock front appears to depend only weakly on the medium despite the fact that phase transitions and shock wave splitting occur in some of these media during the insensitive interval. The radius of the shock front in one medium approaches and then deviates from that in another gradually, so that the insensitive interval is not sharply defined.

Although the reason for the existence of an insensitive interval for this collection of media is currently not well understood from a fundamental physical point of view, it appears to stem from a particular correlation among the relevant properties of these media. It is known that the relevant properties of other geologic media are *not* correlated in this way, so that the radius of the shock front in these media during the "insensitive interval" is very different from the radius of the shock front in the silicates cited above. Thus, the existence of an insensitive interval must be established by test experience or modeling, and is only assured for certain geologic media.

In using insensitive interval scaling, the shock wave sensing cable must be placed close enough to the center of the explosion that it samples the insensitive interval. Yield estimates are then derived by fitting a simple empirical formula, called the Los Alamos Formula, to the RVT data in this interval.⁴⁸ If the radius R of the shock front is expressed in meters, this formula is

$$R(t) = a W^{1/3} (t/W^{1/3})^b, \quad (16)$$

where W is the yield of the explosion in kilotons, t is the elapsed time since the beginning of the explosion in milliseconds, and a and b are constants. According to the assumption on which the algorithm is based, the values of a and b do not depend on the medium (because of this, eq. [16] has sometimes been referred to as the Los Alamos "universal formula"). The values of the constants a and b are typically determined by fitting equation (16) to a selected interval of RVT data from a collection of nuclear explosions. Different individuals and groups have found different values of a and b at various times. Even the values used by a single group have changed with time by amounts that have caused estimated yields to change by tens of percent. The values of a and b used here are 6.29 and 0.475.⁴⁸

The Los Alamos Formula is a simple power law that approximates the actual RVT curve during the insensitive interval. This is illustrated in Figure 4, which compares the Formula with a model of the evolution of the shock wave produced in granite by a spherically-symmetric point explosion with a yield of 62 kt. In practice, the Los Alamos Formula is usually first fit to a broad interval of radius vs. time data that is thought to include the insensitive interval. The result is a sequence of yield estimates. Due to the departure of the Formula from the actual RVT curve at both early and late times, the sequence of yield estimates typically forms a U-shaped curve. This is illustrated in Figure 5, which shows the sequence of yield estimates obtained by applying the Formula to the relatively high-quality SLIFER data from the *Piledriver* explosion in granite (yield estimation using CORTEX data cannot be illustrated here because at present all CORTEX data remain classified). If the assumptions on which the algorithm is based are satisfied, the yield estimates near the bottom of the curve should approximate the actual yield of the explosion.

In the usual form of the algorithm, only the RVT data that fall within a certain predetermined interval chosen on the basis of previous experience (the so-called *algorithmic interval*) are actually used to make the final yield estimate. Because both the location and the extent of the algorithmic interval depend on the yield W of the explosion (both are proportional to $W^{1/3}$), an iterative procedure must be followed in estimating the yield of an explosion whose yield is initially unknown. Table 3 lists the algorithmic interval for several yields.

The sensitivity of an individual yield estimate to an error in the inferred location of the shock front depends on the position of the data point within the algorithmic interval and the yield of the explosion. For example, the sensitivity

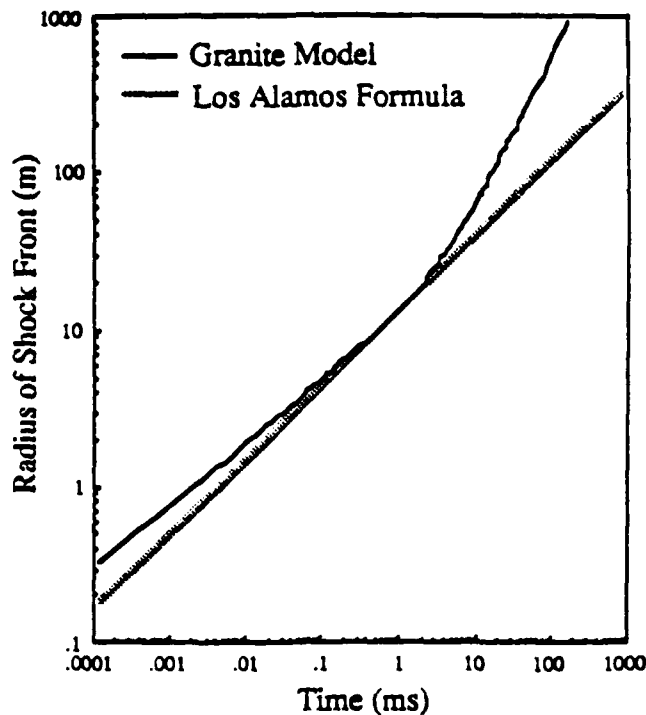


Fig. 4.—Comparison of the Los Alamos Formula with a model of the evolution of a shock wave in granite produced by a spherically-symmetric point explosion with a yield of 62 kt, showing the agreement during the transition interval and the departure of the Formula from the model at earlier and later times. The effect of the phase transition at 400 kbar has been included in the model.

varies from 13 kt m^{-1} at the beginning of the interval to 7.4 kt m^{-1} at the end of the interval, for a 10 kt explosion, and from 77 kt m^{-1} to 18 kt m^{-1} , for a 150 kt explosion.

The algorithmic interval for a yield of 62 kt is indicated in Figure 5 by the two vertical bars at the bottom of the figure. In this example the assumptions on which the algorithm is based appear to be satisfied and the average of the yield estimates that lie within the algorithmic interval is very close to the announced yield of 62 kt.

The insensitiv interval algorithm does not work as well if the assumptions on which it is based are not satisfied. This is illustrated in Figure 6, which shows the yield estimates obtained by fitting the Los Alamos Formula to good-quality SLIFER data from a typical low-yield explosion in alluvium. In this example the RVT data have been scaled in the manner described below in the subsection on similar explosion scaling, so that the derived yield should be 1 kt (the actual yield is classified). The yield estimates given by the Los Alamos Formula are systematically low, ranging from 30% to 82% of the actual yield, and do not form a U-shaped curve. The average of the yield estimates that lie within the algorithmic interval is about 60% of the actual yield. The overall appearance of the yield vs. time curve shows that the assumptions of the algorithm are not satisfied.

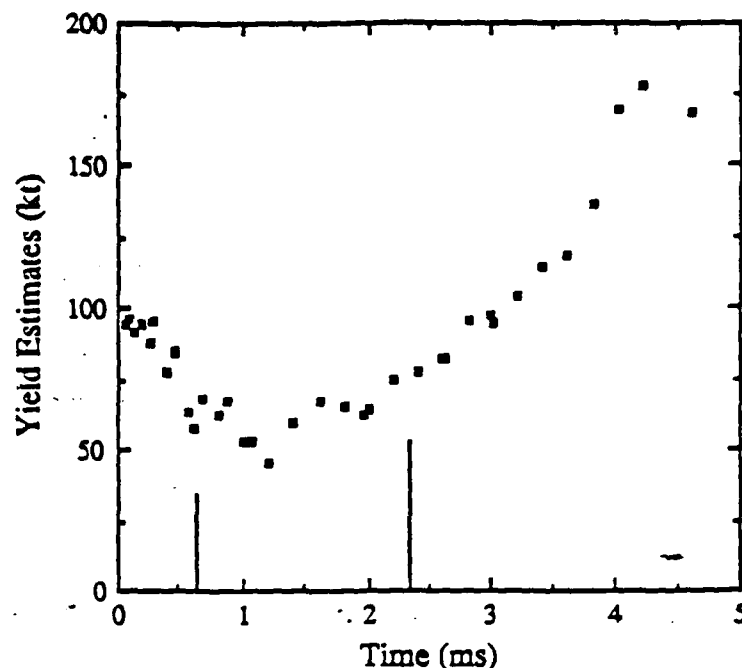


Fig. 5.—Yield estimates derived by applying the Los Alamos Formula to SLIFER data from the *Piledriver* explosion in granite, which had a nominal yield of 62 kt. The so-called algorithmic time interval is the interval between the two vertical bars. The average of the yield estimates within the algorithmic interval is in good agreement with the nominal yield. Note the departure of the yield estimates from 50–70 kt at earlier and later times, which is caused by the deviation of the Formula from the actual radius vs. time behavior of the shock wave outside the algorithmic interval (cf. Fig. 4).

TABLE 3
Algorithmic Intervals^a

Yield (kt)	Time Interval (ms)	Radius Interval (m)
1	0.10 — 0.5	2.1 — 4.5
10	0.21 — 1.1	4.5 — 9.7
50	0.37 — 1.8	7.7 — 17
100	0.46 — 2.3	9.8 — 21
150	0.53 — 2.7	11 — 24

^aThe intervals used by various individuals and groups vary. Throughout this article the algorithmic interval is taken to be from $0.1 W^{1/3}$ ms to $0.5 W^{1/3}$ ms after the beginning of the explosion, where W is the yield of the explosion in kilotons.

A common misconception has been that the algorithmic interval lies within the strong shock region and that the relative insensitivity of yield estimates to the properties of the medium stems from this.⁴⁹ This misconception apparently

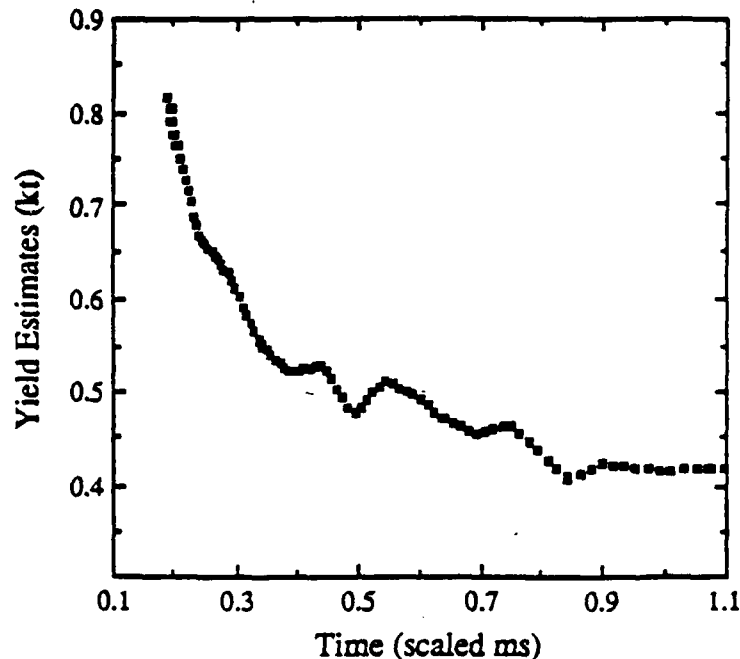


Fig. 6.—Yield estimates derived by applying the Los Alamos Formula to SLIFER data from a low-yield explosion in alluvium (note that the vertical axis is offset from zero). The data have been scaled so that the estimated yield should be 1 kt. Difficulties in applying the insensitive interval algorithm to low-yield explosions in alluvium are not uncommon.

has arisen at least in part because the interval formerly used to estimate the yields of nuclear explosions in the atmosphere using hydrodynamic methods is within the strong shock region. As explained earlier, the relative insensitivity of the radius of the shock front to the medium in the algorithmic interval would be explained if the algorithmic interval were within the strong shock region and if the motion were self-similar. The formula for the radius of the shock front would then be a power-law function of time like the Los Alamos Formula, except that the exponent of t would be 0.4. However, in reality the shock wave motion is not self-similar during the algorithmic interval, for current test geometries and the yields permitted by the TTBT. In fact, the shock wave is not even strong during this interval, since the shock speed is only a few times the low-pressure plastic wave speed while the peak pressure is much less than the pressure required to achieve the limiting density ratio. Indeed, the exponent of time usually used in the Los Alamos Formula, 0.475, is significantly greater than the exponent 0.4 characteristic of a strong, self-similar shock wave. The power-law Los Alamos Formula is not a theoretical result like the power-law formula for the radius of a strong, self-similar shock wave. Rather, it is an approximate, empirical relation, which was obtained by fitting a simple power-law expression to a selected interval of RVT data from a collection of nuclear explosions in several different media.

B. Similar Explosion Scaling

As long as the final state of the shocked rock is independent of the rate of compression and explosions take place in the same medium, the RVT curve depends only on the yield of the explosion.¹¹ If $R = f(t)$ is the RVT curve for an explosion with a yield of 1 kt, then the curve for a yield of W kt is

$$R = W^{1/3} f(t/W^{1/3}). \quad (17)$$

The model given by equation (11), for example, scales with yield in this way. Thus, if the shock wave motion is measured for a given explosion of unknown yield that takes place in the same medium as a previous explosion, and if both RVT data and the yield are available for the previous explosion, the yield of the given explosion can be estimated by comparing the two sets of RVT data. This algorithm is called similar explosion scaling.

Similar explosion scaling can utilize data outside the insensitive interval and works well if the ambient media at the two explosion sites are sufficiently similar. However, in practice it has sometimes proved difficult to ascertain whether the relevant properties of the media are similar enough to give the desired accuracy. As a result, application of this algorithm has sometimes led to unexpected errors in the estimated yield.

If hydrodynamic yield estimation is adopted for TTBT verification, similar explosion scaling would presumably be used where possible to check the results of insensitive interval scaling.

C. Semi-Analytical Modeling

Another approach that is useful for studying the evolution of shock waves in geologic media and for estimating yields is semi-analytical modeling.^{30,32,50-52} In this approach both the properties of the ambient medium and the motion of the shock front are treated in a simplified way that nevertheless includes the most important effects. The result is a relatively simple, semi-analytical expression for the radius of the shock front as a function of time. If the required properties of the ambient medium are known and inserted in this expression, the yield of an explosion can be estimated by fitting the expression to RVT data.

Semi-analytical algorithms can make use of data over a more extensive interval than the interval used in the insensitive interval algorithm. For example, application of equation (10) to particle velocity data taken 8 m from an explosion with a nominal yield of 10.4 kt gave an estimated yield of 10.3 kt.⁵³ Even data taken at a radius of 13.5 m, well outside the algorithmic interval, gave a yield of 7.9 kt. Semi-analytical models can also be used to estimate the uncertainty in the yield caused by uncertainties in the properties of the ambient medium.

D. Numerical Modeling

If a treatment that includes the details of the equation of state and other properties of the ambient medium is required, if it is desired to utilize RVT

data in the region where the shock wave has split, or if the explosion is not spherically symmetric, modeling of the motion of the shock front using numerical codes may be necessary.⁵⁴ In principle, such simulations can provide RVT curves that extend over much of the shock wave evolution, making it possible to base yield estimates not only on data from the transition interval but also on data from later phases of the shock wave evolution. In these later phases, the shock wave is no longer hydrodynamic, so yield estimation methods that make use of such data are not hydrodynamic methods. Yield estimates made using data from later phases are fairly sensitive to the equation of state and constitutive relations that characterize the ambient rock. Since these are accurately known only for a few rocks, accurate yield estimates are possible using late-time data only for explosions in a few geologic media.

E. Complications

In order to use hydrodynamic methods, RVT data must be taken within meters of the center of the explosion (cf. Tables 2 and 3). At such small distances, the arrangement of the nuclear charge and the canister or canisters containing it and the diagnostic equipment can introduce asymmetries in the expansion of the shock wave that will affect the yield estimate:

- A long canister can produce a shock wave that is initially cylindrical and hence crushes the sensing cable in such a way that only part of the total yield is sensed over most of the interval sampled, as shown schematically in Figure 7a. The physical size of canisters and diagnostic lines-of-sight tend to pose more of a problem for tests of nuclear directed-energy weapons than for tests of traditional nuclear weapons.⁵⁵
- Explosions of nuclear charges in tunnels may be accompanied by complicated (and unanticipated) energy flows and complex shock wave patterns. If significant energy reaches the sensing cable ahead of the ground shock and shorts or destroys it before the ground shock arrives, the CORTEX data will describe that flow of energy and not the motion of the ground shock. Moreover, the ground shock itself may become sufficiently distorted that the RVT data are confusing or misleading.

In complicated geometries, disturbing effects like these are difficult to analyze and correct for using data from a single sensing cable, since it senses only the depth of the shallowest point where a pressure wave first crushes it, at a single azimuth and lateral displacement from the explosion. As a result, unambiguous interpretation of the data may become difficult or impossible. The disturbing effect of a canister and emplacement hole of given size is less for higher-yield than for lower-yield explosions, since the hydrodynamic region extends further from the canister and emplacement hole for a higher-yield explosion. In addition, higher-yield charges are not usually exploded in tunnels.

In addition to the potential disturbing effects of the nuclear test design, any errors in characterizing the surrounding geologic media will introduce errors

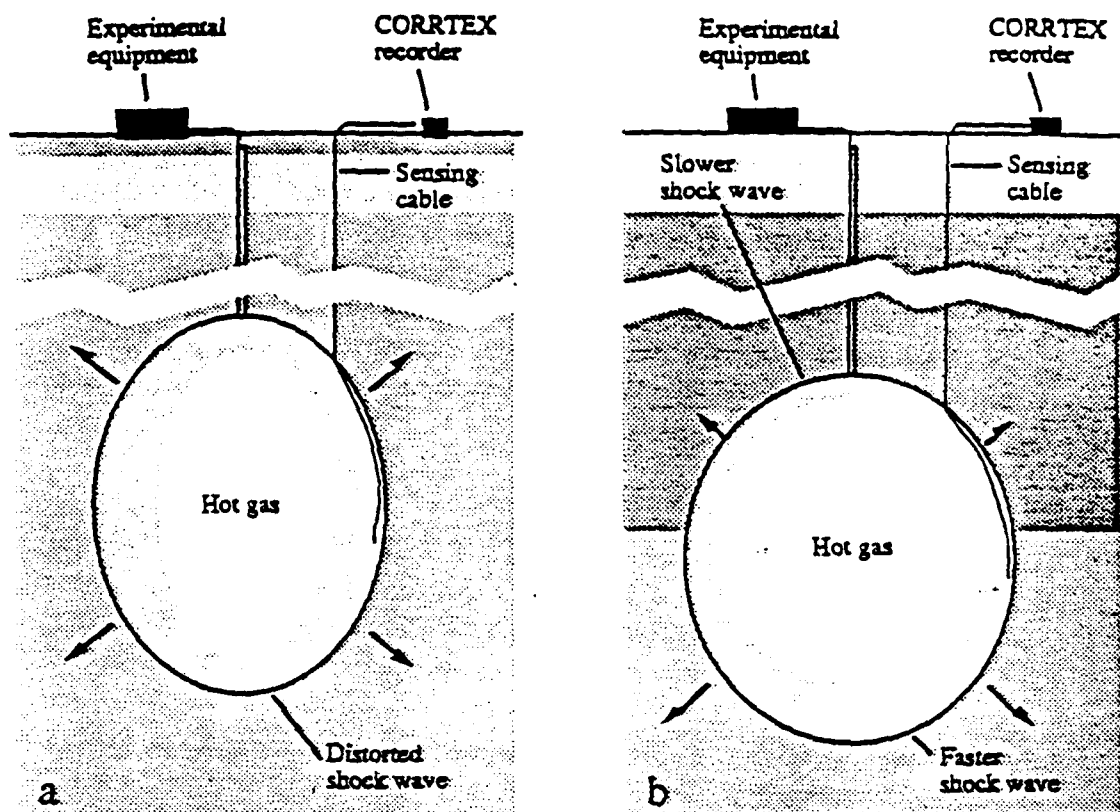


Fig. 7.—Schematic drawings illustrating disturbance of hydrodynamic yield estimates by (a) a nuclear test design that produces a cylindrical shock wave during the algorithmic interval and (b) slowing and weakening of the upward traveling portion of the shock wave when it encounters a much denser geologic stratum.

in yield estimates derived using hydrodynamic methods. Incorrect assumptions about the average properties of the rock surrounding the emplacement hole may bias the yield estimate, decreasing its accuracy, while small-scale variations in the rock will cause scatter in the RVT record, decreasing the precision of the yield estimate. In addition, geologic structures like that shown in Figure 7b can affect the yield estimate. At the Nevada Test Site, for example, the alluvial deposits are weakly consolidated erosion products of the surrounding mountains with physical properties that vary widely. Layers of gravel, the residues of ancient stream beds, are often encountered in drilled holes. While most RVT records follow the expected behavior, an occasional event will produce an RVT record whose irregular behavior defies simple explanation. Such behavior has been attributed to spatial variations in the ambient medium.²⁶

F. Summary of Yield Estimation

Shock waves produced by an underground nuclear explosion propagate differently in different media and different geologic structures. As a result, knowledge of the rock and geologic structure within $\sim 10(W/1\text{ kt})^{1/3}$ meters of the center of the explosion is required in order to make accurate yield estimates using hydrodynamic methods. The evolution of the shock wave can also be af-

fects by the test geometry. Knowledge of the test geometry and/or limitations on the aspects that could disturb the RVT data are therefore required in order to make accurate yield estimates. Several different yield estimation algorithms have been developed. Like the algorithms used to estimate yields from seismic data, these algorithms involve some complexity and require sophistication to understand and apply correctly.

Hydrodynamic yield estimation methods have not yet been studied as thoroughly or as widely as seismic methods. Although approximately 100 tests have been carried out with the CORTEX sensing cable in the emplacement hole, only a few have been carried out with the cable in a satellite hole^{5,6} (in addition, SLIFER data from sensing cables in satellite holes are available for several tens of earlier explosions). Moreover, no systematic and comprehensive review of the scientific evidence concerning the accuracy and precision of U.S. hydrodynamic yield estimation methods has yet been carried out. The U.S. Department of State has asserted that the methods are accurate to within 15% (at the 95% confidence level) of radiochemical yield estimates for tests with yields greater than 50 kt in the geologic media in which tests have been conducted at the Nevada Test Site.^{5,6} According to these same reports, the algorithm is expected to have an uncertainty of a factor of 1.3 at the 95% confidence level if used in a treaty-monitoring context at the Soviet test sites near Shagan River to monitor explosions with yields greater than 50 kt.^{5,6} However, some scientists familiar with the methods believe that the uncertainty could be somewhat larger.⁴

Some key terms that have been introduced in this discussion are listed and explained in Table 4.

VI. APPLICATION TO MONITORING TEST BAN TREATIES

A. Assuring Accuracy

Ambient medium.—As explained in §III and §V, the physical properties and geologic structure of the media around the emplacement point affect the evolution of the shock wave produced by an underground nuclear explosion. Thus, it is important in test ban monitoring to gather information about the types of rock present at the test site and their properties, including their chemical composition, bulk density, porosity, gas-filled porosity, and degree of liquid saturation, as well as the speed of sound in the rock and any specific features of the local geologic structure that could affect the yield estimate. Such data are easily gathered by the party conducting the explosion. Availability of the required data to the monitoring party could be assured by appropriate negotiated cooperative measures.

Some information about the geologic medium at the test site can be obtained by examining the contents of the hole drilled for the CORTEX sensing cable. Verification could be improved by cooperative arrangements that would also allow observation of the construction of the emplacement hole, removal and examination of rock core or rock fragments from the wall of the emplacement

TABLE 4
Glossary of Hydrodynamic Yield Estimation Terms

Term	Explanation
Algorithmic interval	The interval, in time usually 0.1–0.5 scaled ms after the beginning of the explosion, from which data is selected in applying insensitive interval scaling
CORRTEX technique	A technique for measuring the position of the shock front expanding away from an underground explosion by determining the round-trip travel time of electrical pulses sent down a sensing cable placed in a hole in the ground near the explosion
Insensitive interval scaling	A yield estimation algorithm in which the Los Alamos Formula is fit to measurements of the position of the expanding shock front as a function of time during the algorithmic interval
Los Alamos Formula	The empirical formula used in insensitive interval scaling to make yield estimates by fitting to shock front radius vs. time data
Plastic wave interval	The interval in radius and time in which the speed of the weakening hydrodynamic shock wave is less than 1.2 times the low-pressure plastic wave speed
Similar explosion scaling	A yield estimation algorithm in which shock front radius data are compared with similar data from a previous explosion in the same medium
SLIFER technique	A technique for measuring the position of the shock front expanding away from an underground explosion by determining the resonant frequency of an electrical circuit that includes a sensing cable placed in a hole in the ground near the explosion
Strong shock interval	The interval in radius and time in which the density just behind the shock front is close to its limiting value
Transition interval	The interval in radius and time in which the density just behind the shock front is less than 80% of its limiting value but the speed of the hydrodynamic shock wave is still greater than 1.2 times the low-pressure plastic wave speed

hole, examination of any logs or drill core from existing exploratory holes, removal and examination of rock core or rock fragments from the walls of existing exploratory holes, and if necessary, construction of new exploratory holes.

There is precedent for such cooperative arrangements in the Protocol to the Peaceful Nuclear Explosions Treaty (PNET), which explicitly established that hydrodynamic methods could be used to monitor the yield of any explosion with a planned aggregate yield greater than 150 kt and which specified verification measures like those described here.³

Test geometry.—As explained in §V, the arrangement of the nuclear charge and the canister or canisters containing it and the diagnostic equipment can introduce asymmetries in the expansion of the shock wave which disturb the yield estimate. In using hydrodynamic methods to estimate the yields of one's own tests, the design and placement of the nuclear charge and related equipment are known and can be taken into account. This information may not be available when monitoring the nuclear tests of another party. Cooperative agreements to ensure satisfactory placement of the sensing cable and to exclude nuclear test geometries that would significantly disturb the yield estimate are therefore required.⁵⁵⁻⁵⁸

Such agreements could, for example, limit the length of the canister containing the nuclear charge and the cross-sectional dimensions of the emplacement hole, and mandate stemming of the emplacement hole with certain types of materials. Such agreements could also provide for observation of the emplacement of the nuclear charge and the backfilling of the emplacement hole, confirmation of the depth of emplacement, and limitations on the placement of cables or other equipment that might interfere with the CORRTEX measurement. For test geometries that include ancillary shafts, drifts, or other cavities, additional measures, such as placement of several sensing cables around the nuclear charge emplacement point, may be required to assure an accurate yield estimate. For tunnel shots, sensing cables could be placed in the tunnel walls or in a special hole drilled toward the tunnel from above. Again, there is precedent for such cooperative measures in the PNET.³

The restrictions on the size of canisters and diagnostic lines-of-sight that would be required even with the sensing cable placed in a satellite hole would cause some interference with the U. S. nuclear testing program at NTS. However, these restrictions have been examined in detail by the U.S. nuclear weapon design laboratories and the Department of Energy, and have been found to be manageable for the weapon tests that are planned for the next several years.^{55,58} For the more distant future, the disadvantages of the required restrictions on testing must be weighed against the potential contributions to treaty monitoring that could be made by hydrodynamic methods.

In summary, the accuracy that could be achieved using hydrodynamic methods to monitor test ban treaties depends on the amount of information about the ambient medium that can be gathered by the monitoring party and the nature and extent of cooperative arrangements to limit disturbing effects. Some

tests and simulations to identify troublesome configurations have been carried out, but such configurations have not been thoroughly explored. Given the possibility that hydrodynamic yield estimation methods may have to be used to monitor treaty compliance in an adversarial atmosphere, the possibility of deliberate efforts to introduce error or ambiguity, and the tendency for worst-case interpretations to prevail, additional research to improve understanding of the accuracy and precision of these methods and to reduce further the chances of confusion, ambiguity, spoofing, or data denial is very important.

B. Minimizing Intrusion

Hydrodynamic yield estimation methods are more intrusive than seismic methods for several reasons:

- Drilling of the emplacement and satellite holes and emplacement of the nuclear charge and diagnostic equipment typically take 8-10 weeks. In order to monitor these operations, personnel from the monitoring country would have to be present at the test site of the testing country for ten weeks or so before as well as during each test, and would therefore have an opportunity to observe test preparations. The presence of these personnel would pose some operational security problems.^{55,58,59}
- The exterior of the canister or canisters containing the nuclear charge and diagnostic equipment must be examined to verify that the restrictions necessary for the yield estimate to be valid are satisfied. For tests of nuclear directed-energy weapons, this examination could reveal sensitive design information unless special procedures are followed.⁵⁸
- Sensing cables and electrical equipment will tend to pick up the electromagnetic pulse (EMP) generated by the explosion. A detailed analysis of the EMP would reveal sensitive information about the design and performance of the nuclear device being tested.⁶⁰

Intrusiveness can be minimized by careful attention to monitoring procedures and equipment. For example, the electrical recording equipment can be designed to avoid measuring sensitive information about the nuclear devices being tested. CORTEX equipment has been designed in this way, and the United States could insist that any Soviet equipment used at the Nevada Test Site be similarly designed. Placement of the recording equipment in Faraday cages and other measures could be mandated in order to minimize the chance that sensitive information concerning the test could be picked up by the monitoring party.

The security problems posed by opportunities to observe test preparations are more severe for nuclear directed-energy weapon tests, since they tend to have larger and more complex diagnostic systems and canister arrangements which, if fully revealed to the Soviets, might disclose sensitive information.^{55,58} The United States has determined that the Soviet personnel and activities that would be required at the Nevada Test Site to monitor U. S. tests would be acceptable both from a security standpoint and from the standpoint of their

effect on the U.S. test program. Detailed operational plans have been developed to accommodate such visits without adverse impact on operations.⁵⁵

C. Specific Applications

Threshold Test Ban Treaty.—As noted earlier, hydrodynamic yield estimation methods have been proposed by the Reagan administration as a new routine measure for monitoring the sizes of nuclear tests, in order to verify compliance with the 150 kt limit of the TTBT. These methods can be used only if appropriate changes in the TTBT are negotiated. In addition, adequate cooperative arrangements would need to be made in order to assure the accuracy of the resulting yield estimates. To reduce the cost and intrusiveness of verification, use of hydrodynamic methods could be restricted to tests with planned yields greater than some threshold that is an appreciable fraction of 150 kt. As an alternative to routine use, hydrodynamic methods could be used to help calibrate seismic yield estimation methods by measuring the yields of one or more nuclear calibration explosions at test sites,^{57,61} provided that the medium in which the explosion takes place is within U.S. test experience and the necessary procedures are followed to assure the accuracy of the hydrodynamic yield estimate.

Peaceful Nuclear Explosions Treaty.—As it stands, the PNET does not provide for use of hydrodynamic yield estimation except for salvos in which the planned aggregate yield is greater than 150 kt.³ Thus, if the TTBT is modified to allow hydrodynamic yield estimation for all weapon tests with planned yields above a certain value, the purpose of the modification could in principle be circumvented by carrying out weapon tests as "peaceful" nuclear explosions of "planned yield" less than or equal to 150 kt, unless the PNET is also modified to close this loophole.

Low-Threshold Test Ban Treaty.—Tamped underground nuclear explosions as small as a few kilotons produce shock waves that evolve in the same way as those produced by explosions of larger yield. However, because the algorithmic interval for such explosions is much closer to the nuclear charge and diagnostic canisters than the algorithmic interval for tests with much larger yields, the disturbing effect of the canisters is generally more serious for such explosions. Moreover, low-yield tests can be and often are set off at shallow depths in softer material, such as alluvium, or in tunnels. As a result, the propagation of the shock wave can differ markedly from the models on which standard hydrodynamic yield estimation methods are based, causing confusion or error in the yield estimate. There can also be significant variations in the motion of the shock wave from explosion to explosion under these conditions.

In addition, serious practical, operational, and engineering problems arise in trying to use hydrodynamic methods to estimate the yields of explosions with yields of a few kilotons. For one thing, the sensing cable must be placed very close to the nuclear charge. Drilling a satellite hole within 2-5 meters of the emplacement hole to the depth at which the nuclear charge is placed (typically ~250 m even for low-yield explosions in vertical shafts), which would be required in order

to use hydrodynamic methods to monitor explosions with yields in the 1-10 kt range, is at or beyond the capabilities of current drilling techniques.⁴¹ The need for such close placement would also necessitate more stringent restrictions on the maximum sizes and orientations of the nuclear charge and the diagnostic canisters. Such restrictions might be deemed an unacceptable interference with test programs. The alternative of using small canisters with numerous diagnostic lines-of-sight to the detonation point could disturb the shock wave and introduce errors in the yield measurements. Finally, because the shock wave radii to be measured are much smaller at low yields, survey errors become much more important.

It is possible that some of these difficulties could be circumvented by developing models and algorithms that would allow routine use of RVT data at radii beyond the hydrodynamic interval, as discussed in §V. Others could be alleviated by conducting tests with simple geometries to define and calibrate test sites so that insensitive interval scaling could be accurately applied. However, these and other potential solutions to the problems that would be encountered in monitoring low-yield tests using hydrodynamic methods have not yet been carefully and thoroughly studied. Thus, at the present time hydrodynamic yield estimation methods could not be used with confidence to monitor compliance with threshold test bans in which the threshold is less than several tens of kilotons.

Comprehensive Test Ban Treaty.—Hydrodynamic methods can only be used with shock wave data taken within meters of the center of the explosion. Moreover, their use requires advance notification of tests, extensive preparations in advance at the test site, and the presence of monitoring personnel during the test. Hydrodynamic methods were not intended and clearly are not able to detect or identify unannounced, remote, or clandestine nuclear tests. As a result, they cannot contribute to monitoring a comprehensive test ban.

VII. DISCUSSION

Hydrodynamic methods could be a potentially valuable component of a cooperative program to improve monitoring of the TTBT and PNET. Hydrodynamic yield estimates are more affected than seismic estimates by local features that disturb the evolution of the shock wave, such as canisters or local geologic structures. For explosions larger than ~10 kt, seismic yield estimates are less affected by local features but are affected by how the shock wave evolves into seismic waves. Thus, yield estimates based on hydrodynamic methods would complement yield estimates based on seismic methods.

Apparently, a systematic and comprehensive review of the scientific evidence concerning the accuracy and precision of hydrodynamic yield estimation methods has not yet been conducted in the United States. Nor has the United States any experience as yet in applying these methods at Soviet test sites. As noted in §V, the U.S. Department of State has asserted that yield estimates based on CORTEX data analyzed using the insensitive interval algorithm are

expected to have an uncertainty of a factor of 1.3 at the 95% confidence level, for explosions larger than 50 kt at the Soviet test sites near Shagan River. If achieved, such an uncertainty would be slightly smaller than the uncertainty of a factor of 1.5-1.6 that according to recent studies⁴ could probably be achieved using the best current seismic methods, if the test site is not calibrated and measurements are made only outside the Soviet Union.

Hydrodynamic methods could also be used to help calibrate seismic methods. According to the results of a recent study, application of the best current seismic methods to measurements made outside the Soviet Union is expected give yield estimates that have an uncertainty of a factor of 1.3 at the 95% confidence level, the same precision reportedly expected for hydrodynamic methods, once the Soviet test site is properly defined and calibrated.⁴

In its first years, the Reagan administration declined to resume the U.S.-Soviet negotiations on a CTBT that had been adjourned in 1980. However, in the summer of 1986 both parties agreed to begin new, low-level talks on nuclear testing.^{60,62} The Soviet Union reportedly pressed for resumption of negotiations on a CTBT and emphasized seismic verification methods whereas the United States sought an agreement allowing use of hydrodynamic methods as an additional verification measure for the TTBT and PNET.⁶² Then, in April 1987, the Soviet Union suggested experiments at the U.S. and Soviet test sites as a way to advance the negotiations and to compare verification methods.⁶³⁻⁶⁵ As a result of further discussions^{66,67} the United States and the Soviet Union announced on September 17, 1987, that they had agreed to renew full-scale negotiations on nuclear testing limitations.^{68,69} Formal negotiations began in November 1987 in Geneva.

The parties agreed as a first step to negotiate verification measures for the TTBT and PNET. In these negotiations, the sides agreed to make no changes in the treaties themselves but to negotiate a new protocol to each treaty, which would specify verification provisions.⁶⁹ Once these two treaties have been ratified, the parties have agreed to begin negotiations on ways to implement a step-by-step program to further limit nuclear tests.⁶⁹

At the Washington Summit in December 1987, President Reagan and General Secretary Gorbachev agreed to conduct a joint verification experiment (JVE) at each other's nuclear test sites.⁶⁹ In this experiment, one nuclear charge with a yield near the 150 kt limit will be exploded at the Nevada Test Site; another will be exploded at the Soviet test site near Semipalatinsk. The purpose of the experiment is to compare the verification techniques advocated by the two sides in the Geneva negotiations.

In preparation for the JVE, 20 nuclear testing experts from each side visited the other side's nuclear testing site in January 1988 to familiarize themselves with how each side conducts tests and to provide a basis for the design and conduct of the JVE.^{69,70} The second round of the current U.S.-Soviet test ban negotiations began in February 1988, and remained in continuous session until June 30, 1988.⁴⁷ In March, drilling of the necessary holes began.⁶⁹

A 200-page agreement specifying the procedures for the JVE was signed on May 31, 1988, as part of the Moscow Summit.⁷¹ The agreement, which with its annex totals nearly 200 pages, specifies that both sides will have the opportunity to measure the yield of each of the JVE explosions using both hydrodynamic methods (with sensing cables in both the emplacement hole and a satellite hole) and teleseismic methods (with recordings made at five designated seismic stations of each side).^{69,71} On June 28, 1988, each side shared with the other seismic data on five high-yield underground tests conducted after January 1, 1978, but before January 1, 1988, recorded at the five designated seismic stations at teleseismic distances from its nuclear test site, as well as associated geological and geophysical information and estimates of the yields of the five events.⁶⁹

Originally, the Soviet Union had suggested that each side furnish the nuclear charge to be exploded at the other side's test site, so that the monitoring side would be able to compare the yields estimated using both seismic and hydrodynamic methods with a yield estimate based on knowledge of the design of the nuclear charge.^{64,65,66} However, the JVE is actually being carried out with a U.S. nuclear charge at the U.S. test site and a Soviet nuclear charge at the Soviet test site.⁷¹ As a result, the monitoring party will not have available an estimate of the yield of the explosion that is independent of both seismic and hydrodynamic methods. In order to solve this calibration problem, the sides have agreed to accept the hydrodynamic yield estimate derived from the data taken with the sensing cable in the emplacement hole as the yield of the explosion for the purposes of the experiment.

The JVE test at the Nevada Test Site, code-named *Kearsarge*, was completed as scheduled on August 17, 1988. The Soviet Union reportedly gathered hydrodynamic data on the explosion using sensing cables in the emplacement and satellite holes, regional seismic data from the seismic stations that have been installed nearby as part of the NRDC-Soviet Academy cooperative verification project,^{72,73} and teleseismic data from its own national seismic network.^{46,72} The Soviets will also be able to use data from a variety of seismographic stations in the United States that routinely publish such data. The United States reportedly gathered hydrodynamic data on *Kearsarge* and will also be able to use data from its national seismic network as well as published data. Data on this explosion was exchanged in Geneva on August 30, 1988, at the start of the third session of the current round of negotiations.

According to press reports following *Kearsarge*,⁷⁴ the nuclear charge was expected to have a yield of about 140 kt. However, one U.S. hydrodynamic measurement reportedly gave a yield estimate slightly more than 150 kt while the other gave a yield estimate slightly more than 160 kt. In contrast, data from the U.S. seismic network reportedly confirmed expectations that the yield of the explosion would be about 140 kt. If the yield of the explosion was 140 kt and if the uncertainty of the hydrodynamic yield estimation algorithm used by the United States is a factor of 1.15 at the 95% confidence level at the Nevada Test Site, as asserted by the U.S. Department of State,^{5,6} then the probability

that a single yield estimate would be larger than 160 kt is only about one in 40, while the joint probability that one yield estimate would be more than 150 kt and the other more than 160 kt is about one in 300. As a result of this outcome, the U.S. Department of Energy has reportedly decided to make a radiochemical yield estimate in order to better assess the yield of *Kearsarge*.

As this is being written, preparations for the explosion at the Semipalatinsk Test Site, which is currently scheduled for September 14, 1988, are in progress. Because of its strong advocacy of hydrodynamic methods, the United States has chosen not to gather regional seismic data on this explosion but will gather hydrodynamic data using CORTEX sensing cables in the emplacement and satellite holes.⁷⁵ The United States will also have available teleseismic data from its national seismic network. Regional seismic data on the test at Semipalatinsk may be recorded by seismic stations that have been installed nearby as part of the NRDC-Soviet Academy cooperative verification project. If it is, this data would presumably also be available to the U.S. government. Thus, the JVE will provide the United States with a variety of information with which to improve its seismic estimates of the yields of past as well as future nuclear tests at Semipalatinsk.

The U.S. drive to negotiate hydrodynamic monitoring of the TTBT and PNET raises several important issues concerning the interaction of technology and public policy. From a purely scientific point of view, gathering of additional data on U.S. and Soviet test yields and achievement of better precision in estimating yields are advances. However, the principal purpose of TTBT and PNET monitoring is presumably to increase the security of the United States and the Soviet Union, rather than to advance scientific studies.

From a security perspective, the purpose of requiring a precision of a factor of 1.3 at the 95% confidence level for monitoring the 150 kt limit of the TTBT appears unclear. Such a precision means that there is only one chance in four that a single explosion with a yield of 164 kt would appear to be in compliance with the TTBT. The administration has not explained why such a high precision is needed, nor has it explained why current seismic methods, which according to recent studies could achieve a similar precision using measurements made only outside the Soviet Union, are still considered inadequate. Was the requirement for such high precision established after a careful review of the purposes of the TTBT and PNET and the costs and benefits of requiring various levels of precision? Or was it adopted simply because it was the highest precision that was thought to be achievable with hydrodynamic methods at the time the policy advocating such methods was adopted?

In assessing whether this degree of precision is necessary, one must bear in mind the fact that if tests producing yields of 150 kt are performed, then approximately half of unbiased yield estimates will be above 150 kt (just as approximately half will be below 150 kt), no matter how precise the estimates. Moreover, both parties to the TTBT recognized at the time it was signed that the yield of a nuclear test cannot always be predicted accurately in advance,

so that even if the planned yield is 150 kt, some variation from this value may occur.³

Whatever the justification for requiring a precision of a factor of 1.3 at the 95% confidence level in monitoring the 150 kt limits of the TTBT and PNET, the considerations that enter an assessment of the precision that would be optimal in monitoring possible future lower-threshold test ban agreements are different, and depend on the size of the threshold. For some choices of threshold, an uncertainty much greater than a factor of 1.3 might be adequate. However, the emphasis on achieving a precision of a factor of 1.3 at 150 kt could establish this precision in the public mind as a requirement that must also be met in monitoring any future limitations on nuclear testing. Such a development could impede progress toward achieving such limitations.

Similarly, the emphasis on hydrodynamic monitoring of the TTBT and PNET might also establish this verification technology as a required part of any low-threshold test ban agreement. Given the difficulties that would have to be overcome to apply these methods with confidence to monitoring low-yield tests, such a requirement could become an additional impediment to achieving a low-threshold test ban.

The current test ban negotiations and the Joint Verification Experiment have led to increased openness about the U.S. and Soviet nuclear testing programs and increased cooperation between the national security establishments of the United States and the Soviet Union. According to participants, both sides have approached the JVE as a useful joint venture, in which it has been possible to make enquiries, receive answers, and resolve concerns.⁶⁹ In addition to these benefits, the JVE will also provide useful data on the seismic properties of the U.S. and Soviet test sites that may help to resolve past concerns over compliance and to advance toward ratification of the TTBT and PNET.

VIII. CONCLUSIONS

Hydrodynamic methods were developed primarily as a tool to estimate the sizes of explosions in the U.S. and Soviet nuclear test programs and are inherently more intrusive than seismic methods. From a scientific point of view hydrodynamic methods are no more "direct" than seismic methods. Nor are hydrodynamic methods necessarily free of systematic error. Nevertheless, the evidence that is available suggests that use of hydrodynamic methods at Soviet test sites is likely to give estimates of the yields of Soviet tests larger than 50 kt that are slightly more precise than could be achieved using the best currently available seismic methods, if the Soviet test sites are not calibrated and seismic data are collected only outside the Soviet Union, and if the cooperative arrangements needed to assure the accuracy of hydrodynamic methods are successfully negotiated and followed. On the other hand, the precision of the best available remote seismic methods is expected to be comparable to the precision that could be achieved using hydrodynamic methods if the Soviet test sites are appropriately defined and calibrated. Significant engineering, operational, and analysis

problems need to be solved before hydrodynamic methods could be used to monitor with confidence a low-threshold test ban. The methods are not relevant to a comprehensive test ban.

Possible advantages of an agreement allowing use of hydrodynamic yield estimation methods for monitoring the TTBT and PNET include the increased openness about nuclear testing programs to which this would lead and the ongoing cooperation between the national security establishments of the United States and the Soviet Union that would occur, as well as the additional deterrence against cheating and the additional assurance of compliance that use of these methods would provide. Such an agreement would also establish the principle of on-site inspection at nuclear test sites, would make available more accurate data on the yield distribution of large nuclear tests, and might indirectly limit nuclear testing in the future.

Possible disadvantages of seeking an agreement allowing routine use of hydrodynamic methods for test ban monitoring include the potential difficulty of successfully negotiating such use, the restrictions that such use would impose on nuclear test designs, which could adversely affect nuclear testing programs in the future, and the operational security problems at test sites that would be created by the continuous presence there of dozens of monitoring personnel from the other country. Adoption of hydrodynamic yield estimation methods for routine test ban monitoring might also create an impediment to further progress in limiting nuclear testing. Many of the possible advantages of hydrodynamic methods could be gained and many of the possible disadvantages of routine use avoided by an agreement allowing limited use of such methods, such as for calibration of seismic methods.

It is a pleasure to thank T. Lomelino for help in analyzing SLIFER data and rock equations of state and I. Fushiki and G. Miller for help in simulating ground shock propagation. I also thank D. Barr, D. Bernstein, S. Drell, D. Eilers, A. Gancarev, R. Geil, D. Hafemeister, J. Hannon, M. Heusinkveld, R. Hill, R. Jeffries, S. Keeny, D. Metzger, W. Moss, M. Nordyke, W. Panofsky, J. Parmentola, P. Richards, R. Scribner, J. Sullivan, G. van der Vink, and D. Westervelt for helpful comments. I am grateful to Professor Sidney Drell and the other staff members for the warm hospitality extended to me at the Center for International Security and Arms Control, Stanford University, where part of this work was completed, and to the University of Illinois Program in Arms Control, Disarmament, and International Security and the Carnegie Corporation of New York for support. Some of the analysis reported here was carried out for the Congressional Office of Technology Assessment study of seismic verification of nuclear testing treaties and for a study of test ban monitoring sponsored by the Institute for Defense Analyses.

NOTES AND REFERENCES

1. Fermi estimated the yield of the first nuclear test *Trinity* within seconds after the explosion, using a simple hydrodynamic technique. See R. Rhodes, *The Making of the Atomic Bomb* (New York: Simon and Schuster, 1988), p. 674. A model of the shock wave produced by a nuclear explosion in air was published by G. I. Taylor, *Proc. Roy. Soc., A* 151, 421 (1950).
2. See, for example, G. W. Johnson, G. H. Higgins, and C. E. Violet, *J. Geophys. Res.*, 64, 1457 (1959) and J. H. Nuckolls Rep. UCRL-5675 (Livermore, Cal.: Lawrence Radiation Laboratory, 1959).
3. *Arms Control and Disarmament Agreements* (Washington, D.C.: U.S. Arms Control and Disarmament Agency, 1982).
4. For an excellent introduction to the scientific and policy issues related to nuclear test ban verification, see U.S. Congress, Office of Technology Assessment, *Seismic Verification of Nuclear Testing Treaties*, OTA-ISC-361 (Washington, D.C.: U.S. Government Printing Office, May 1988).
5. U.S. Department of State, Bureau of Public Affairs, *U.S. Policy Regarding Limitations on Nuclear Testing* (U.S. Department of State Special Report No. 150, August, 1986).
6. U.S. Department of State, Bureau of Public Affairs, *Verifying Nuclear Testing Limitations: Possible U.S.-Soviet Cooperation* (U.S. Department of State Special Report No. 152, August 14, 1986).
7. The uncertainty at the 95% confidence level is usually cited simply because it is conventional to cite the uncertainty at this confidence level in seismology. It is important to note, however, that this confidence level may not be appropriate in the context of test ban monitoring and that the uncertainty is quite sensitive to the choice of confidence level. For example, for a normal distribution a factor of two uncertainty at the 95% confidence level is the same as a factor of 1.27 uncertainty at the 50% confidence level. This means that there is only one chance in four that a single explosion with a yield of 190 kt would appear to have a most likely yield of 150 kt or less and only one chance in 16 that two such explosions would appear to be treaty compliant.
8. In order to prevent seepage of radioactive gases to the surface, the depth of burial (DOB) must be at least $120(W/1 \text{ kt})^{1/3}$ m. This requires a DOB of at least 650 m for a 150 kt explosion. When the DOB given by this relation (the so-called scaled depth of burial, or SDOB) would be relatively small, and in media with a substantial water content, the actual DOB is increased in order to achieve containment of radioactive gases. Thus, the actual DOB of an explosion at the Nevada Test Site is normally not less than 250 m. See S. Glasstone and P. J. Dolan, *The Effects of Nuclear Weapons* (Washington, D.C.: U.S. Government Printing Office, 1977), p. 261.
9. For an example of a stemming plan, see H. D. Glenn, T. F. Stubbs, and J. A. Kalinowski, and E. C. Woodward, Rep. UCRL-89410 (Livermore, Cal.: Lawrence Livermore National Laboratory, August 1983). For a shorter account,

- see H. D. Glenn, T. F. Stubbs, and J. A. Kalinowski, in *Shock Waves in Condensed Matter*, ed. Y. M. Gupta (New York: Plenum, 1986), p. 639.
10. S. Glasstone and P. J. Dolan, *op. cit.*, pp. 61-63 and 260-262.
 11. H. L. Brode, *Ann. Rev. Nuclear Sci.*, 18, 153 (1968).
 12. R. W. Terhune, H. D. Glenn, D. E. Burton, H. L. McKague, and J. T. Rambo, *Nucl. Tech.*, 46, 159 (1979).
 13. W.K.H. Panofsky, in *Threshold Test Ban Treaty and Peaceful Nuclear Explosions Treaty, Hearings Before the Committee on Foreign Relations, United States Senate* (Washington, D.C.: U. S. Government Printing Office, 1987), at pp. 66 and 247.
 14. P. G. Richards, this volume.
 15. L. R. Sykes, J. F. Evernden, and I. Cifuentes, in *Physics, Technology, and the Nuclear Arms Race*, AIP Conf. Proc. No. 104 (New York: American Institute of Physics, 1983).
 16. W. J. Hannon, *Science*, 227, 251 (1985).
 17. J. F. Evernden and C. B. Archambeau, in *Arms Control Verification* (New York: Pergamon-Brassey's, 1986).
 18. J. F. Evernden, C. B. Archambeau, and E. Cranswick, *Rev. Geophys.*, 24, 143 (1986).
 19. R. B. Barker, in *Threshold Test Ban Treaty and Peaceful Nuclear Explosions Treaty, Hearings Before the Committee on Foreign Relations, United States Senate* (Washington, D.C.: U. S. Government Printing Office, 1987), at pp. 8, 19, and 89-90.
 20. D. A. Vesser, in *Threshold Test Ban Treaty and Peaceful Nuclear Explosions Treaty, Hearings Before the Committee on Foreign Relations, United States Senate* (Washington, D.C.: U. S. Government Printing Office, 1987), at p. 94.
 21. S. R. Foley, in *Threshold Test Ban Treaty and Peaceful Nuclear Explosions Treaty, Hearings Before the Committee on Foreign Relations, United States Senate* (Washington, D.C.: U. S. Government Printing Office, 1987), at p. 11.
 22. J. H. McNally, in *Threshold Test Ban Treaty and Peaceful Nuclear Explosions Treaty, Hearings Before the Committee on Foreign Relations, United States Senate* (Washington, D.C.: U. S. Government Printing Office, 1987), at pp. 27 and 99-101.
 23. H. A. Holmes, in *Threshold Test Ban Treaty and Peaceful Nuclear Explosions Treaty, Hearings Before the Committee on Foreign Relations, United States Senate* (Washington, D.C.: U. S. Government Printing Office, 1987), at pp. 5 and 108.
 24. See Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Phenomena* (New York: Academic Press, 1967 [English Translation]), pp. 685-705.
 25. Ya. B. Zel'dovich and Yu. P. Raizer, *op. cit.*, pp. 741-746.
 26. F. Holzer, *J. Geophys. Res.*, 70, 893 (1965).
 27. Ya. B. Zel'dovich and Yu. P. Raizer, *op. cit.*, pp. 49-50 and 705-710.
 28. W. R. Perret and R. C. Bass, Rep. SAND 74-0252 (Albuquerque, N.M.: Sandia

- Corp., 1975).
29. J. D. Johnson and S. P. Lyon, Rep. LA-10391-MS (Los Alamos, N.M.: Los Alamos National Laboratory, 1985).
 30. W. C. Moss, *J. Appl. Phys.*, **63**, 4771 (1988).
 31. J. D. Johnson, unpublished.
 32. F. K. Lamb, ACDIS WP-2-87-2 (Urbana, Ill.: University of Illinois Program in Arms Control, Disarmament, and International Security, February 1987).
 33. One can also obtain a closed expression for $R(t)$ for any Hugoniot $H(u)$ that is piece-wise linear.
 34. Ya. B. Zel'dovich and Yu. P. Raizer, *op. cit.*, Chap. I and XII.
 35. L. I. Sedov, *Similarity and Dimensional Methods in Mechanics* (New York: Academic Press, 1959 [English Translation]).
 36. G. Barenblatt, *Similarity, Self-Similarity, and Intermediate Asymptotics* (New York: Consultants Bureau, 1979 [English Translation]).
 37. See, for example, Fig. 2 of R. C. Bass, in *Shock Waves in Condensed Matter*, ed. Y. M. Gupta (New York: Plenum, 1986), p. 633.
 38. G. C. Schmitt and R. D. Dick, Rep. LA-UR-85-231 (Los Alamos, N.M.: Los Alamos National Laboratory, 1985).
 39. C. F. Virchow, G. E. Conrad, D. M. Holt, and E. K. Hodson, *Rev. Sci. Instrum.*, **51**, 642 (1980).
 40. R. G. Deupree, D. D. Eilers, T. O. McKown, and W. H. Storey, Rep. LA-UR-80-3382 (Los Alamos, N.M.: Los Alamos National Laboratory, 1980).
 41. During preparations for the 1988 U.S.-Soviet Joint Verification Experiment discussed in §VII, the Soviets stated that they did not have the technology to drill satellite and emplacement holes to the required depth (presumably ~ 650 m) while maintaining a horizontal displacement within the tolerance (presumably ~ 10 m) required by the U.S. As a result, the United States flew its drill rig and crew to the Soviet test site. See C. P. Robinson, Testimony before the Subcommittee on Arms Control, International Security, and Science, House Committee on Foreign Affairs, June 28, 1988.
 42. M. Heusinkveld and F. Holzer, *Rev. Sci. Instrum.*, **35**, 1105 (1964).
 43. W. H. Storey, D. D. Eilers, T. O. McKown, D. M. Holt, and G. C. Conrad Rep. LA-UR-82-558 (Los Alamos, N.M.: Los Alamos National Laboratory, 1982).
 44. Public information sheet on CORTEX (Los Alamos, N.M.: Los Alamos National Laboratory, April 1986).
 45. R. B. Barker, in *Threshold Test Ban Treaty and Peaceful Nuclear Explosions Treaty, Hearings Before the Committee on Foreign Relations, United States Senate* (Washington, D.C.: U. S. Government Printing Office, 1987), at p. 16.
 46. S. Blakeslee, *New York Times*, August 15, 1988, p. A4.
 47. Robert C. Toth, *Los Angeles Times*, March 23, 1988.
 48. R. C. Bass and G. E. Larsen, Rep. SAND 77-0402 (Albuquerque, N.M.: Sandia National Laboratories, 1977).
 49. For example, according to refs. 5 and 6 "The accuracy of the method is believed to be relatively, but not wholly, independent of the geologic medium, provided

the satellite hole measurements are made in the 'strong shock' region ..."

50. M. Heusinkveld, *J. Geophys. Res.*, **87**, 1891 (1982).
51. M. Heusinkveld, *Rep. UCRL-52648* (Livermore, Cal.: Lawrence Livermore National Laboratory, 1979).
52. R. A. Axford and D. D. Holm, *Proc. Nuclear Explosion Design Conference* (Los Alamos, N.M.: Los Alamos National Laboratory, October 1987).
53. W. C. Moss, private communication (1987).
54. For a list of some of the numerical codes currently in use, see F. K. Lamb, "Monitoring Yields of Underground Nuclear Tests", in *Threshold Test Ban Treaty and Peaceful Nuclear Explosions Treaty, Hearings Before the Committee on Foreign Relations, United States Senate* (Washington, D.C.: U. S. Government Printing Office, 1987), p. 359.
55. Sylvester R. Foley, Jr., Assistant Secretary for Defense Programs, Department of Energy, letter to Edward J. Markey, Congressman from Massachusetts (March 23, 1987).
56. S. S. Hecker, in *Threshold Test Ban Treaty and Peaceful Nuclear Explosions Treaty, Hearings Before the Committee on Foreign Relations, United States Senate* (Washington, D.C.: U. S. Government Printing Office, 1987), at pp. 50-54 and 233-235.
57. M. D. Nordyke, in *Threshold Test Ban Treaty and Peaceful Nuclear Explosions Treaty, Hearings Before the Committee on Foreign Relations, United States Senate* (Washington, D.C.: U. S. Government Printing Office, 1987), at pp. 70-71 and 283-284.
58. R. E. Batzel, in *Threshold Test Ban Treaty and Peaceful Nuclear Explosions Treaty, Hearings Before the Committee on Foreign Relations, United States Senate* (Washington, D.C.: U. S. Government Printing Office, 1987), at pp. 56 and 221-224.
59. Seven C5 military transport aircraft were required to ferry drilling equipment and personnel to the Soviet test site for the 1988 Joint Verification Experiment. See C. P. Robinson, Testimony before the Subcommittee on Arms Control, International Security, and Science, House Committee on Foreign Affairs, June 28, 1988.
60. See R. J. Smith, *Washington Post*, September 4, 1986.
61. It has been suggested that the nuclear calibration charges to be detonated at the test sites could be provided by the monitoring country. If they were, a hydrodynamic estimate of the yield might not be needed, since the yields of certain types of nuclear charges are accurately reproducible. Knowledge of the surrounding rock and geologic structure would still be needed to provide assurance that the coupling of the explosion to seismic waves was understood. However, some way would have to be found to provide assurance that no sensitive information about the design of the nuclear weapons of either country would be revealed.
62. M. R. Gordon, *New York Times*, July 10, 1986, p. 1.
63. M. R. Gordon, *New York Times*, April 7, 1987, p. 1.

64. M. R. Gordon, *New York Times*, April 18, 1987, p. 1.
65. W. Pincus, *Washington Post*, April 18, 1987, p. 1.
66. M. R. Gordon, *New York Times*, June 4, 1987, p. 3.
67. A. Rosenthal, *New York Times*, September 1, 1987, p. 3.
68. W. F. Burns, Testimony before the Subcommittee on Arms Control, International Security, and Science, House Committee on Foreign Affairs, June 28, 1988.
69. C. P. Robinson, Testimony before the Subcommittee on Arms Control, International Security, and Science, House Committee on Foreign Affairs, June 28, 1988.
70. Associated Press, *Los Angeles Times*, January 16, 1988.
71. *Agreement Between the United States of America and the Union of Soviet Socialist Republics on the Conduct of a Joint Verification Experiment* (Washington, D.C.: U.S. Government Printing Office, 1988).
72. S. Blakeslee, *New York Times*, August 18, 1988, p. A3.
73. W. Sweet, *Physics Today*, November 1987, p. 83.
74. M. R. Gordon, *New York Times*, September 11, 1988, p. 9.
75. R. B. Barker, News Conference in Geneva, Switzerland, November 20, 1987 (see *Arms Control Today*, December, 1987, p. 26).

B. OPPORTUNITIES FOR LONG-WAVELENGTH IR DETECTORS

Katherine T. Faber,^a Nathan S. Lewis,^b Thomas F. Rosenbaum,^c
Steven J. Sibener,^c and R. Stanley Williams^d

Defense Science Study Group
Institute for Defense Analyses
1801 N. Beauregard Street
Alexandria, VA 22311

ABSTRACT

This document is a brief overview of some recent advances in materials and technology for long-wavelength (8-14 μm) infrared detectors. There have been several breakthroughs in the past five years in the areas of new superconductors, band-structure engineering, device concepts, and refrigeration that require consideration as potential competitors to mercury-cadmium telluride (MCT) devices or for special-niche applications. In addition, the possibility of utilizing extrinsic III-V compound semiconductors for long-wavelength IR detectors has apparently been overlooked, but is shown here to be promising. This report has been written for internal use by the Defense Advanced Research Projects Agency by members of the Defense Science Study Group, which is sponsored by the Institute for Defense Analyses.

^a Dept. of Materials Science, Northwestern University, Evanston, IL 60201

^b Dept. of Chemistry, California Institute of Technology, Pasadena, CA 91125

^c James Franck Institute, University of Chicago, Chicago, IL 60637

^d Dept. of Chemistry & Biochemistry, U C L A, Los Angeles, CA 90024

OUTLINE

I. INTRODUCTION.....	III-49
II. DEVICE CATEGORIES.....	III-50
A. Bolometers.....	III-50
B. Photodetectors.....	III-52
1. Materials Considerations.....	III-52
a. Intrinsic Semiconductors ..	III-52
b. Lattice-Matched Quantum Wells.....	III-53
c. Strained-Layer Superlattices ..	III-56
d. Extrinsic Semiconductors ..	III-58
i. Qualitative considerations.....	III-59
ii. Quantitative analysis.....	III-61
2. Technological Considerations.....	III-67
a. Blocked-Impurity Band Detectors.....	III-67
b. Refrigeration.....	III-69
i. Gas-cycle cooling systems.....	III-70
ii. Magnetic cooling systems.....	III-70
III. CONCLUSIONS.....	III-71
IV. BIBLIOGRAPHY.....	III-72
Appendix.....	III-75
Tables	III-76
Figures	III-79

I. INTRODUCTION

This document was motivated by the authors' interest in focal plane arrays for imaging sources of 8-14 μm IR radiation, which is an atmospheric transmission window, and their concern over the expense and difficulty of material preparation and device fabrication involving the $\text{Hg}_x\text{Cd}_{1-x}\text{Te}$ (MCT) alloy system. It is not intended as an exhaustive review or proposal, but rather as an overview of some new developments that have arisen in the past five years or so that could enable the production of long wavelength IR detectors without resorting to MCT. The authors may have overlooked some areas in which important advances have occurred, and the coverage of the areas reflects our backgrounds and biases. The authors have not seen the topics presented here collected together in any single article or book, and we suggest some specific materials and device structures that may or may not be original ideas. Thus, we view this paper as a useful compendium of research opportunities available to develop alternatives to MCT devices as long wavelength IR sensors and arrays.

The two broad categories of devices that are appropriate for IR sensors are bolometers and photodetectors: the appearance of the high T_c superconductors has warranted a reexamination of the role of the former, while new materials and device concepts have provided a large number of possibilities for the latter. In particular, photodiode arrays based on the maturing technology of the III-V compound semiconductors appear to be especially attractive. Also, new concepts in device design, such as improved avalanche photodetectors, are available to enhance the properties of any given materials system. Finally, developments in refrigeration may become the enabling technology that leads to a new generation of IR sensors. When operated at temperatures below 20K, extrinsically doped Si detectors can outperform MCT detectors at 77K (see Fig. 1), but at present the premium paid in obtaining such low temperatures is considered to be too great. However, there are materials that may be able to outperform MCT at temperatures between 20K and 77K, and thus provide a cost-effective alternative in some applications.

II. DEVICE CATEGORIES

A. Bolometer Detectors

Bolometers are essentially temperature-dependent resistors that can detect optical radiation down to extremely long wavelengths, well into the millimeter range. They are the most sensitive infrared detectors available when cooled to, or below, liquid helium temperatures. Within the constellation of detector technologies, they offer superb detectivity and broad spectral response. It is for these reasons that bolometers are the most widely used detectors in modern infrared astronomy experiments. One should note, however, that important trade-offs exist between sensitivity and response time for these detectors, with typical response times falling in the millisecond range. Recent innovations have led to the fabrication of practical *bolometer arrays*, as described below.

The operation of a bolometer detector is simple to review. The active sensing element is a small chip of a material having a high-temperature coefficient of resistance, typically Ga-doped Ge (Low, 1961), Si, or a superconducting material (Martin and Bloor, 1961) operating at its critical temperature. This element is attached via thermally conducting leads to a stable cryogenic heat sink. The chip itself is usually connected to a bridge or voltage-dividing circuit operating at a given bias current. When optical radiation is absorbed by the bolometer element, a small temperature jump is induced in the device, which causes its resistance to change, and a voltage (after suitable amplification) to be detected. The magnitude of the temperature jump is a function of the bolometer's heat capacity. The response time, τ , is simply given by the ratio of the heat capacity of the element, C , to the thermal conductance, G , between the device and the heat bath: $\tau = C/G$. Note that increasing the thermal conductance has the effect of lowering the response time at the expense of decreased sensitivity. Note, also, that lower heat capacity bolometers are desirable, ultimately favoring Si elements over the more traditional Ga-doped Ge. Background radiation considerations also influence device optimization: G must be large enough to maintain the bolometer at the desired temperature. The applied bias current, which determines the magnitude of the voltage response, cannot be made arbitrarily large due to heating of the device

and the associated increase in noise. High-impedance and low-noise amplifiers operating at 77K are needed to realize optimal performance. Typical performance parameters of a Ga-doped Ge bolometer operating at 2K are: $NEP \sim 10^{-14} \text{ W Hz}^{-1/2}$ and $\tau \sim 10 \text{ msec}$. Cooling to temperatures below 2K allows NEP values in the $10^{-15} \text{ W Hz}^{-1/2}$ region to be accessed. Response times on the order of 1-10 μsec can be achieved when suitable trade-offs are made against detector response.

Based upon the above discussion, one might conclude that superconducting bolometers operating at their transition temperature would offer significant advantages over Ge- or Si-based devices due to the large resistance increases that occur as these materials change to their normal state. Unfortunately, extraneous noise associated with material inhomogeneities and instabilities during the phase transition limits this advantage. Further technical problems arise due to the extreme temperature stability requirements needed to reach optical performance. High- T_c materials do not offer major advantages here, since the thermal noise and fluctuations present at *ca.* 100K set the fundamental limitations on device performance.

The newest generation of bolometers are being fabricated from Si substrates, using ion-implanted metallic (*i.e.*, *low-noise*) contacts. Si has a significantly lower specific heat than Ga-doped Ge, leading to improved response. More important, fabrication techniques can be based on relatively mature Si-processing technologies. *Bolometer arrays* with (1×8) and (2×10) active elements have already been fabricated with excellent results (Low, *et al.*, 1987). The NEP of a (2×10) array operating at 0.34K, achieved with a closed-cycle He refrigerator (Low, *et al.*, 1987), was recently reported to be $4.1 \times 10^{-16} \text{ W Hz}^{-1/2}$ with elements of dimension 24×10 mil. Array response speeds were sufficient to achieve image-refresh rates of 20-40 Hz. High manufacturing yields and device uniformity associated with Si-based technology suggest that larger arrays can certainly be constructed if desired.

To summarize, this section has discussed bolometer detectors and arrays that can currently be fabricated for detecting extremely weak infrared signatures in a low background environment. Drawbacks associated with these devices include extreme cooling requirements, $T \leq 4\text{K}$, and

relatively slow response time. When operational requirements can be met, such as infrared astronomy applications, bolometers are clearly the detector of choice due to excellent broadband response, low noise, and high sensitivity.

B. Photodetectors

Descriptions of the generic types of semiconductor photodetectors have been given by Sze (1981). The basic detection process depends upon the absorption of photons to produce free charge carriers, which can be monitored as a current or voltage. For an intrinsic semiconductor, both an electron and a hole are generated by photoexcitation across the bandgap. Either a free electron or a hole is generated in an extrinsic material by photoexcitation from a donor to the conduction band or from the valence band to an acceptor. Thus, the limitation on which photon wavelengths can be detected is primarily a materials problem involving the availability of semiconductors with the appropriate bandgaps or dopant levels (see Table I for bandgaps and Fig. 2b for the known dopant levels in Ge, Si, and GaAs, both from Sze (1981)).

There are many different types of devices that may be fabricated, once an appropriate material has been selected. The configuration chosen will depend on the desired detection characteristics, such as gain of the device or response time. The primary device categories are photoconductors and photodiodes, with the latter including *p-n* junction, *p-i-n* junction, Schottky Barrier detectors, avalanche diodes, and photocapacitors (MIS devices). Of primary interest here are those systems that can be utilized in focal plane arrays for imaging purposes (Botts, 1988).

1. Materials Considerations

a) Intrinsic Semiconductors

Clearly the most straightforward way of producing an IR detector is with an intrinsic semiconductor with a band gap that corresponds to the photon wavelength that one wishes to detect. Unfortunately, nature has not supplied an element or simple binary compound with an absorption edge in the 8-14 μm range. Thus, the search for an appropriate material turned to ternary compounds, and alloys of the zero bandgap semimetal HgTe and the 1.56 eV bandgap semiconductor CdTe have attracted the most attention. These two binary compounds are

isostructural (see Table II) and infinitely soluble in one another, so that it is possible to vary the bandgap of the alloy across the entire IR range of the spectrum. This property was considered especially attractive, since in principle only one materials technology had to be developed to cover all the IR wavelengths of interest. Unfortunately, the bandgap of the alloy is strongly dependent on the composition, especially for LWIR material, and producing large areas of uniform stoichiometry is extremely difficult in this materials system, especially in a production environment. This is the primary limitation on present MCT technology.

Two other known ternary alloys have bandgaps that approach the desired 10 μm region. The two zinc-blende compounds InSb and InAs also form a continuous range of solid solutions. The $\text{InAs}_{1-x}\text{Sb}_x$ ternary alloy has the property that the 77K bandgap has a local minimum at $x=0.6$ of 0.17 eV (Coderre and Wooley, 1968, see Fig. 3a). This corresponds to a wavelength of 7.3 μm , but as will be seen later the absorbing wavelength can be increased by applying a strain to the intrinsic material. The other ternary is an alloy of InSb and InBi. In this case, the two binaries are not isostructural, since InBi has a tetragonal crystal structure. Thus, the equilibrium solubility of InBi in InSb is only 2.2% (Fig. 3b from Jean-Louis and Hamon, 1969), although with molecular beam epitaxy (MBE) it is possible to obtain a high-quality 3% metastable alloy with an absorption edge of 9.5 μm at 77 K (Noreika, *et al.*, 1982). Although both of these intrinsic materials come close to overlapping the spectral region of interest, neither can span the window and thus by themselves are not broadly applicable.

There are an enormous number of inorganic ternary and quaternary alloys and organic polymers that have not had their optical properties investigated in any depth. These systems provide ample subjects for fundamental studies of material properties and, given the large number of possibilities, there are likely to be several other intrinsic materials with a direct bandgap in the 10 μm region that could prove useful as IR detectors.

b) Lattice-Matched Quantum Wells

The electrical and optical properties of $\text{GaAs}/\text{Al}_x\text{Ga}_{1-x}\text{As}$ quantum wells have been the subject of intense investigation for about a decade. By confining thin films of the narrower

bandgap material (GaAs) between buffer layers of a wider bandgap material (AlAs or $\text{Al}_x\text{Ga}_{1-x}\text{As}$), discrete states resembling those of a particle-in-a-box (PB) are induced in both the valence and conduction bands of the GaAs. Most optical studies of these states to date have been concerned with the analysis of excitonic and band-to-band excitations involving transitions across the GaAs band gap. However, a recent study by West and Eglash (1985) demonstrated that transitions between two different PB subband states in the GaAs conduction band could be observed.

These researchers grew 50 period GaAs/ $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ superlattices with the buffer layers containing the *n*-type dopant Si (a modulation-doped superlattice). Since the lowest PB-like state in the GaAs conduction band is lower than the Si donor state in the $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$, electrons flow from the buffer layers to occupy this subband. Eglash and West observed that their structures had an extremely large-dipole infrared transition that corresponded to the energy separation between the lowest two PB-like states in the conduction band. The peak absorption energies were 152 and 121 meV for GaAs quantum well widths of 65 and 82 Å, respectively. The energy separation between the subbands is determined by the selection of the GaAs layer thickness and the composition of the buffer alloy, which controls the potential barrier height at the edges of the quantum wells. Thus, it is in principle possible to fabricate detectors sensitive to the entire IR spectrum from GaAs quantum wells.

There have been several striking advances in the use of GaAs/AlAs superlattices as single-element IR detectors reported by a group at AT&T Bell Laboratories in the past three years. Their first achievement was to obtain a 95% absorption of 8.2 μm radiation with a single 50 period superlattice by fabricating a waveguide and trapping the radiation in the device by multiple internal reflections (Levine, *et al.*, 1987a). They then demonstrated that by applying a bias voltage perpendicular to the superlattice planes, the mean free path of the photoexcited carriers, and thus the photocurrent, could be enhanced by a process of resonant tunneling through the buffer layers (Levine *et al.*, 1987b, Choi *et al.*, 1987). At a wavelength of 10.8 μm, they obtained a detector responsivity of 1.9 A/W. By pushing to even higher bias voltages, they

discovered that there was an operating regime in which an avalanche gain mechanism (as illustrated in Fig. 4a) increased the responsivity of their detector to 7 A/W at around 80 K, which compares favorably with state-of-the-art MCT photodetectors. Another effect of applying a bias voltage, which may be either an advantage or a nuisance, is that the PB energy levels and the absorption maxima are Stark shifted (Ahn and Chuang, 1987a and 1987b). Thus, it may be possible to construct a tuneable detector, or improve signal to noise in a particular spectral region by modulating the detection wavelength about a signature feature. Of course, responsivity is only one of the criteria to be met in selecting a detector/amplifier system. Other aspects of the device, especially noise limitations, must be considered carefully before any practical system could be designed using these or any other new materials.

Lattice-matched superlattices can be used to make very efficient detectors for any wavelength IR radiation, and the current level of GaAs materials and device technology should enable the construction of sensors that are large in area, uniform in their responsivity and wavelength response, and comparatively inexpensive. The primary limitation at present is how to efficiently couple the radiation into an array of intersubband transition devices. The selection rules for the transitions require a component of the polarization vector *perpendicular* to the plane of the quantum wells (West and Eglash, 1985). Normally incident radiation cannot be absorbed by intersubband detectors, which means that they will be difficult to apply to focal plane arrays unless some means of coupling the radiation into the individual elements can be found.

One possible method of doing this would be to incline a detector array by Brewster's angle with respect to the direction of the incoming radiation (see Fig. 5). The scene of interest would be focused by the appropriate optics onto one end of a dense fiber optic bundle or a channel array cut such that the focal plane would be perpendicular to the light path. The other end of the bundle or array would be cut at an angle of 73° with respect to the light path, and the detector would be mounted flush with this surface. This would enable the focal plane to be projected onto the inclined detector with a minimum of image distortion because of the depth of field limitations of the lenses. The detector would be rectangular, with the dimensions of the elements

designed to compensate for the foreshortening of the image on the inclined array. For example, a typical element size might be $50 \times 170 \mu\text{m}$, with a 128×128 array measuring $1 \times 3.4 \text{ cm}$. The main problems with this scheme are that a fraction of the radiation will be lost in the light guide, and the *s*-polarized component incident on the detector will still not be absorbed, so over 50% of the incident radiation is lost.

Another difficulty is that even when radiation is incident at Brewster's angle, it will be refracted such that inside the semiconductor the angle between the *p*-polarized component of the radiation and the plane of the quantum wells will be small. One could manufacture the superlattice to be thick enough to absorb most of the photons, but this is an expensive and inelegant solution. A better approach might be to allow the radiation to enter the detector from the substrate side, and use the ohmic contact on the surface of each element of the detector array as a reflecting mirror. This will maximize the intensity of the electric field in the superlattice and allow more efficient coupling of the radiation into the absorbing region. Other schemes for such a coupling involving diffraction gratings are also possible, but would be more difficult and expensive to fabricate.

c) Strained-Layer Superlattices

Another method of engineering the band gap of materials is by introducing lattice strain. Although this could be accomplished by subjecting a material to an externally applied load, it can also be achieved by forming a layered structure of two lattice-mismatched materials. Within the plane, layers must be under tension or compression to maintain coherency with the material above or below. In-plane tension or compression results in a Poisson compression or tension, respectively, out of the plane of the lattice (see Fig. 6). Such strains result in the distortion of a cubic lattice to one having tetragonal symmetry.

Both the tetragonal distortions and the in-plane strains in the lattice are known to shift, either increase or decrease, the semiconductor band gap by 10 to 100's of meV (Osborn, 1986a). A variety of systems have been examined: GaAs/GaAsP and GaAs/GaInAs strained-

layer superlattices (SLS) are known to have increased band gaps, whereas GaP/GaAsP, GaSb/AlSb, GaAs/InAlAs, and Si/SiGe have reduced band gaps (Osbourne, 1986b).

Perhaps most promising for focal plane arrays are materials in the InAsSb system (Osbourne, 1984). Unstrained InSb has a band gap, $E_g = 0.17$ eV at 0 K, while the composition InAs_{0.39}Sb_{0.61} exhibits a minimum in bandgap (0.11 eV at 0 K) across the InAs-InSb binary, as noted earlier. Alternating layers of InSb and InAs_{0.39}Sb_{0.61} to produce tensile strains in the InAsSb layer exceeding 0.4% were predicted to be sufficient to reduce the bandgap such that absorption at 12 microns was feasible at 77K (Osbourne, 1984). The first results for SLS structures of InAs_{0.06}Sb_{0.94}/InAs_{0.3}Sb_{0.7} grown by MBE were reported early last year (Kurtz, *et al.*, 1988) and demonstrated absorption in the 8 to 12 μ regime. Another candidate for SLS's for far infrared detectors is InSb/InSbBi. As noted earlier, unstrained InSb_{0.97}Bi_{0.03} has an absorption edge of 9.5 μ and a larger lattice constant than InSb.

One of the major difficulties in the fabrication of such devices is the prevention of misfit defects (dislocations) at layer interfaces. A new model details conditions for the nucleation and generation of dislocations at the layer interface (Hirth and Evans, 1986). The nucleation stress decreases linearly with layer spacing and lattice mismatch. The critical layer thickness for the GaAs/GaAs_{0.5}P_{0.5} system has been measured to be 35 nm; calculations using the Hirth-Evans model are in reasonable agreement (39 nm). Hence, design of defect-free SLS's is relatively straightforward by means of the stress analysis to determine the maximum layer spacing for a given composition.

A second consideration in the development of these new materials is in the absorption coefficient. Although projected absorption coefficients are high, little actual data has been reported in the literature. A high absorption was reported in InAs_{0.06}Sb_{0.94}/InAs_{0.3}Sb_{0.7} (Kurtz, *et al.*, 1988), but it is uncertain if the absorption was influenced by surface reflection, impurity absorption or defect scattering.

The potential of III-V SLS's for focal plane arrays is quite promising since the bandgaps are predicted to be less sensitive to compositional variation than Hg_{1-x}Cd_xTe (Fig. 8). The III-V

materials considered for these strained-layer superlattices are inherently more chemically stable than the HgCdTe compounds, and mechanically more robust.

d) Extrinsic Semiconductors

As a major phase of our work in novel concepts for infrared materials, we have investigated the potential new applications for extrinsic compound semiconductors as infrared detectors. Currently, extrinsic Si devices are used in some IR detection applications (Sclar, 1976); thus improved extrinsic materials may play a significant role in further advancing IR detector technology. It has been shown that at low background levels, superior signal-to-noise performance should be displayed by cooled extrinsic detectors as opposed to intrinsic materials (Blouke, *et al.*, 1973). Thus, for the highest performance devices applied for sighting against "cold" backgrounds (such as outer space, etc.), extrinsics should be quite attractive infrared detector materials. For more conventional applications, the impetus toward use of extrinsic devices largely derives from the fact that Si-based detectors can be monolithically combined with Si amplifiers to produce an integrated detector system. This contrasts with existing HgCdTe technology, which requires "bump-bonding" methods to attach the HgCdTe detector array to a Si-based MUX amplifier. In addition to the high level of complexity of such a device, the difference in thermal expansion coefficients between the HgCdTe detector material and the Si amplifier causes reliability problems in the completed units, especially after repeated thermal cycling. Furthermore, the fabrication of acceptable FLIR modules from HgCdTe currently suffers from nonuniformity in wavelength response of the different detector elements in the array, as mentioned earlier.

Alternative extrinsic materials may permit advances in ultimate signal-to-noise performance and eliminate processing difficulties in current infrared detector applications. We have thus explored the possibilities for extrinsic detector combinations that might utilize current technology to fabricate new materials systems and/or to obtain new dopants in existing materials. In general, extrinsic materials have the disadvantages smaller absorption coefficients and lower temperature requirements than intrinsic detector materials (Long, 1967; Borrello, *et al.*, 1971; Beyen, *et al.*,

1966). Therefore, at this stage of development, the use of novel extrinsics should be viewed as a complement to existing extrinsic materials technology rather as direct competition with current intrinsic HgCdTe technology. Of course, improvements may be expected in the performance of extrinsic compound semiconductors as the learning curve is traced in their research, development and production stages, and these devices may then look even more attractive than at present. Within this framework, we have developed some strategies for exploring novel extrinsic detector materials. Several candidate systems are described below, and a representative example is analyzed quantitatively in order to provide a comparison with existing extrinsic Si detector technology.

i. Qualitative considerations

A key consideration for any detector system is the optical response threshold of the absorber. For Si, no known dopant level is at the correct energy (approximately 0.1 eV) to obtain optimal response in the important 8-12 micron near infrared region, as shown in Fig. 2b (Sze, 1981; Soref, 1967). Thus, there should be other materials that are better suited for use as extrinsic detectors for this wavelength range. However, the driving force to use Si extrinsic detectors has arisen largely from the ease of device fabrication using Si technology.

In the recent past, III-V compound-semiconductor fabrication methods have made great advances relative to their rather primitive, troublesome beginnings. The use of extrinsic GaAs should be especially attractive for future generation long-wavelength IR detectors. This is based on several advantageous optoelectronic properties of the III-V based materials: 1) the well-documented speed advantages of amplifiers and optoelectronic devices fabricated from III-V's, 2) advances in III-V device technologies to make monolithic processing a reality, and 3) control over band gaps and other materials properties by use of alloy composition methods. Additionally, the advances in epitaxial growth techniques (MBE, metal-organic vapor phase epitaxy (MOCVD), etc.) have made the uniform growth of complicated alloy structures quite feasible on a routine basis.

A search of the literature reveals that there are several possibilities for extrinsic detectors that utilize GaAs and other III-V materials. A particular example, to be discussed as a case study in this work, is the Mn/GaAs system, because enough is known about it to make quantitative predictions. Mn is a known *p*-type dopant in GaAs (Ilegems, 1975); it is quite soluble; it has the correct dopant energy level for response in the 8-12 micron range (trap energy=0.106 eV from electrical measurements (Sasaki, *et al.*, 1985)), and it does not introduce deep level traps that would reduce the lifetime of the GaAs sample. Curiously, other than one incomplete Soviet report (Antonov, 1981), to our knowledge there is essentially no work described in the open literature on the fabrication and characterization of IR detectors made from extrinsic III-V materials. As will be shown below, such systems do have a great deal of promise.

Other dopant/semiconductor combinations are undoubtedly also available in the III-V materials system, especially if the advanced epitaxial growth methods are employed to create alloys of a desired composition. Along these lines, a recent study indicates that the energy level of a dopant relative to the band edge of the semiconductor can be varied by changing the alloy composition in the $Al_{1-x}Ga_xAs$ series (Bantien and Weber, 1987); thus, dopants which have possibly been excluded from consideration to date, due to nonoptimal binding energies in pure GaAs, might well be quite satisfactory if incorporated into an alloy of $Al_{1-x}Ga_xAs$, $GaAs_{1-x}P_x$, $In_{1-x}Ga_xAs$, or other similar systems. There has been very little systematic study of dopant levels in the III-V alloy systems to date. We suggest that these possibilities be surveyed and explored in some depth in order to identify suitable dopant/alloy candidates for extrinsic detectors in various wavelength regions.

Of course, a suitable optical threshold for the 8-12 micron region is only one of several requirements that needs to be met in order to achieve satisfactory detector performance. Below, we have quantitatively analyzed the possible performance of the Mn/GaAs system, in order to obtain an estimate of the performance of this type of III-V based extrinsic detector. We find that Mn/GaAs is quite competitive with existing Si extrinsics based on signal-to-noise and detectivity criteria, and potentially offer further advantages in terms of the speed of electronic amplifiers

fabricated from the III-V semiconductors. At present, not enough information is available in the open literature to determine whether the crucial systems parameters of optimized III-V extrinsic detectors can be competitive with existing HgCdTe performance; however, in view of the advances in technology and the relative ease of control over device uniformity and electrical properties in III-V systems, these materials should be worthy of substantial attention as potential new detector systems.

ii. Quantitative analysis of Mn/GaAs detector performance

The key figure of merit for an infrared detector is the spectral detectivity, D^* . In order to estimate the D^* attainable for the proposed extrinsic materials, it is necessary to consider the noise mechanisms that contribute to the performance of a particular detector. The general equation (Long, 1980) for D^* is as follows:

$$D_{\lambda}^* = \frac{\eta \lambda}{U^{1/2} hc [\eta \Phi_b + I_d(V) / (qG^2A) + kT / (q^2G^2RA)]^{1/2}} \quad (1)$$

where $U=4$ for a photoconductive detector (as is the case for most applications with extrinsic materials), and $I_d(V)$ is a current function that expresses the recombination/generation noise of the material/dopant combination (*vide infra*). Other quantities in this expression are λ (the photon wavelength), η (the quantum yield), h (Planck's constant), c (speed of light), Φ_b (background photon flux), k (Boltzmann's constant), T (detector temperature), q (electron charge), G (gain of device), R (effective resistance of the device), and A (detector element area). As a benchmark for comparison, current intrinsic HgCdTe detectors operating in the 8-12 micron region (22% Cd; 78% Hg) exhibit D^* values of 5×10^{11} (cm-Hz^{1/2}/W) at 77 K. In contrast, extrinsic Si detectors in this same spectral range (e.g., Si:As) require substantial additional cooling, and typically exhibit D^* values of $< 10^7$ (cm-Hz^{1/2}/W) at 77 K and 1×10^{10} (cm-Hz^{1/2}/W) at ≈ 30 K (Soref, 1967).

Another figure of merit often used for photovoltaic-type detectors is the R_0A product (where R_0 is the resistance at zero bias and A is the detector area). The D^* at negligible background radiation and zero bias voltage is related to this figure of merit as follows:

$$D_{\lambda}^* = \frac{q\eta\lambda(R_0A)^{1/2}}{2hc(kT)^{1/2}} \quad (2)$$

Thus, a D^* of 4.85×10^{11} (cm-Hz^{1/2}/W) for HgCdTe corresponds to an R_0A product of 10 ohm-cm² for $\lambda_c = 12.4$ μ m and $T = 77$ K. This value is representative of current state-of-the-art performance for intrinsic HgCdTe materials reported in the open literature as of 1988.

For comparison, the background-limited infrared performance (BLIP) value of D^* for any type of detector is given by the following expression:

$$D_{\lambda}^* = \frac{\eta\lambda}{U^{1/2} hc [\eta\Phi_b]^{1/2}} \quad (3)$$

where $U=2$ for a photovoltaic element and as above $U=4$ for a photoconductive element. This equation assumes a 2π field of view aperture into a blackbody source supplying Φ_b of background photons resulting from thermal fluctuations of the source. Thus, using the background photon flux of 9×10^{17} photons/(sec \cdot cm²) appropriate for $\lambda_c=12.4$ μ m and a 300 K blackbody source, we find that the BLIP $D^* \approx 4.51 \times 10^{10}$ (cm-Hz^{1/2}/W) for photovoltaic type systems and the BLIP $D^* = 3.18 \times 10^{10}$ (cm-Hz^{1/2}/W) for photoconductive systems.

This photovoltaic-mode BLIP D^* corresponds to a R_0A product of ~ 0.3 ohm \cdot cm²; values of R_0A substantially in excess of this number would not appreciably increase the signal-to-noise performance for a detector utilizing a constant 2π field of view onto a 300 K background. Of course, smaller angle field-of-view applications would require higher R_0A products for BLIP, as would viewing into colder background environments (Long, 1980). Thus, the D^* values obtained to date with isolated HgCdTe elements exceed these 2π field of view, 77 K

requirements; as mentioned previously, however, fabrication of arrays with uniform response and performance has proven to be quite difficult. In part, this occurs because the tunability of the HgCdTe bandgap with changes in alloy composition makes control over the stoichiometry such a crucial and difficult feature of an optimal array fabrication process. Such difficulties might be avoidable with the use of a fixed threshold system, such as an extrinsic device.

The key to the performance of extrinsic detectors is the second term in Equation (1); the strong temperature dependence of $I_d(V)$ dictates the maximum operating temperature that can be used to allow D^* for the device of interest to be larger than $D^*(\text{BLIP})$. Usually, the gain G of a photoconductive type detector can be made large enough so that the third term in Equation (1) is negligible relative to the second term (Long, 1980); thus, the calculation of D^* at any temperature reduces (to first order) to obtaining a value for $I_d(V)$.

In extrinsic materials, the dominant source of noise is recombination/generation noise due to the majority carriers. A simplified explanation of this noise is as follows: when the photoconductor is exposed to a dark voltage V , this produces a corresponding current I . Any fluctuations in the majority carrier density will contribute to fluctuations in the conductance of the sample, which will produce differing values of I for the fixed value of V . These fluctuations will then define the lower limit on the signal that can be detected and resolved from the majority carrier noise.

The equation defining such recombination/generation noise for an extrinsic p-type semiconductor is as follows (Van Vliet, 1967):

$$g_{th} = \frac{p_0 A t}{\tau_p} , \quad (4a)$$

where

$$\tau_p = \frac{1}{B(N_A - N_D)} , \quad (4b)$$

p_0 is the density of equilibrium majority carriers (holes, for our p-type example), p_1 is defined as $p_1 = N_V \exp[(E_{vb} - E_t)/kT]$ with E_t the dopant level relative to the valence band energy, N_V is the

density of states in the valence band, t is the thickness of the detector element, τ_p is the majority carrier lifetime, and N_A and N_D are the acceptor and donor densities, respectively. The key quantity that is not under experimental control is the coefficient B , which expresses the capture cross section of majority carriers by the dopant. Applying the theorem of microscopic reversibility, this coefficient also therefore characterizes the thermal emission rate from the dopant level into the majority carrier band.

Substituting into Equation (4) allows us to obtain an expression for the second term in Equation (1), as follows:

$$I_d(V) / (qG^2A) = B(N_A - N_D)p_1t . \quad (5)$$

This quantity must then be less than $\eta\Phi_b$ for the detector to exhibit background limited performance.

Another important parameter is the thickness of the device. Extrinsic materials are usually weak absorbers of the photons that excite carriers from or to donors or acceptors because of the low dopant concentration. Hence, these devices generally need to be much thicker than the corresponding intrinsic materials of the same threshold wavelength. This increased thickness in turn contributes to excess noise (or equivalently, more severe temperature requirements in order to obtain suitably low noise values). Quantitatively, the constraint for absorption of photons is:

$$\alpha = \sigma_A(N_A - N_D) , \quad (6)$$

where σ_A is the absorption cross section and α is the absorption coefficient. This implies that $(2.5)\alpha^{-1}$ cm of material are required in order to absorb 90% of the incident photons at the desired wavelengths.

Note that in Equation (5), the product of $(N_A - N_D)$ and t is found; thus the device thickness can be minimized if the dopant is extremely soluble in the material. Solubility limits of the dopant therefore place a lower limit on the thickness of the detector. However, at very high dopant levels, an insulator-conductor Mott transition can be reached, and this can limit extrinsic

semiconductor device fabrication to dopant concentrations well below the solid solubility. Moreover, the threshold wavelength often tends to increase at very high dopant concentrations, because of broadening of the dopant band states. Again, this will set an upper limit on the accessible dopant density and a lower limit on the thickness of the device. These considerations are specific to each material/dopant system, and we can treat only certain examples for which the relevant data are available from the literature with the precision necessary to perform the desired analysis for a general D^* estimate.

For Mn in GaAs, very high ($>10^{19} \text{ cm}^{-3}$) solubilities have been observed (Woodley, 1973) and the Mott transition does not occur until dopant densities of $\approx 10^{19} \text{ cm}^{-3}$ have been reached. Thus, these two factors do not severely limit the Mn/GaAs. Above dopant densities of $2.5 \times 10^{18} \text{ cm}^{-3}$, the absorption edge starts to move to lower energy, and this is therefore the limiting factor for the dopant concentration in the example system (Brown, 1972). The absorption cross section σ_A at $12 \mu\text{m}$ of Mn in GaAs has been reported to be 10^{-16} cm^2 ; thus Equation (6) implies that $50 \mu\text{m}$ -thick samples (with 2 optical passes) are needed to absorb essentially all of the light in the $8\text{-}12 \mu\text{m}$ region. This is the first constraint that is necessary in order to proceed with the analysis.

In order to obtain the other values needed in the analysis, further information on the quantities in Equation (5) is required for the specific materials in question. Values for p_1 have been obtained using the expression $p_1 = N_v \exp[(E_{vb} - E_t)/kT]$, where $E_{vb} - E_t = 0.11 \text{ eV}$ for Mn in GaAs, and the most reliable value for N_v of GaAs has been taken from the literature (Blakemore, 1982). This yields $p_1 = 7.2 \times 10^{10} \text{ cm}^{-3}$ at 77 K .

Using these values in Equation (5) and the recently reported value of B for the Mn/GaAs system of $3.1 \times 10^{-9} \text{ cm}^3/\text{sec}$ (Montelius, *et al.*, 1988), we calculate that the crucial second term in the denominator of Equation (1) has the value $2.8 \times 10^{22} \text{ m}^{-2}\cdot\text{s}^{-1}$. This value in Equation (1) implies that D^* at 77 K equals $1.9 \times 10^{10} (\text{cm}\cdot\text{Hz}^{1/2}/\text{W})$ (without the background photon flux contribution). This is within a factor of 2 of the BLIP D^* for photoconductive-type detectors with a 2π field of view and a 300 K background.

For comparison, D^* values of currently available extrinsic Si detectors are typically $<10^7$ ($\text{cm-Hz}^{1/2}/\text{W}$) at 77 K. These extrinsic detectors require cooling to 25-30 K to achieve background-limited detection values of D^* into a 300 K blackbody (Pines and Baron, 1974). Thus, the Mn/GaAs system would appear to exceed the performance of currently available Si extrinsic detectors on the basis of their expected signal-to-noise performance. The dopant level is closer to the desired cut-off wavelength than the available Si dopants; the dopant energy can be tuned by choice of the proper III-V alloy (Bantien and Weber, 1987), and there are other advantages involved with the high speed and superior optoelectronic properties of the III-V materials. There may be other factors that would limit the Mn/GaAs system, but the above analysis shows that further study of extrinsic III-V's is warranted.

Further consideration of the key parameters in the above equations suggests that there is no fundamental physical reason why other extrinsic materials cannot exhibit superior detectivity to that estimated for the Mn/GaAs system. In principle, the limits for recombination/generation noise of a 0.11 eV trap level would be very similar to those for a 0.11 eV direct (intrinsic) gap transition, provided that the dopant level is optimized for high optical absorption and a low value of B (low majority carrier capture cross section). Values of B vary greatly for different trap levels, depending on whether the trap is coulombically attractive, repulsive or neutral (Lang and Logan, 1975; Mitonneau, *et al.*, 1977); also, trap binding energies can be manipulated by proper choice of alloy composition within the III-V materials family. A survey of these possibilities might turn up more promising dopant/absorber combinations for use in the desired 8-12 μm region of the infrared, in that further reductions in B for an extrinsic III-V would yield performance comparable to the intrinsic HgCdTe system as presently produced.

An additional feature worthy of note is that the Mn/GaAs detectors can be made to exhibit D^* values of $>10^{11}$ ($\text{cm-Hz}^{1/2}/\text{W}$) merely by cooling the devices to 45-50 K. At these temperatures, they should be competitive with extrinsic Si detectors at 20K, and should also exhibit D^* values at 50K higher than those for HgCdTe detectors at 77 K. Thus, another alternative to obtaining improved IR detector arrays might involve trading the ease of fabrication

of extrinsic III-V materials against incurring some additional cost to cool the devices to temperatures below 77 K.

In summary, the above analysis shows that at least one of the III-V family of extrinsic materials should provide superior electronic and signal-to-noise performance to existing extrinsic Si detectors at a particular operating temperature or, alternatively, provide the same performance at much higher operating temperatures. The detectivity theoretically achievable for the Mn/GaAs system operating at 77 K is estimated to be within a factor of 2 of the background-limited value for a 2π field of view element with a 300 K background temperature. It is clearly conceivable that alternative dopants may exist for GaAs or other III-V materials that yield background limited performance in a variety of applications. The use of III-V alloys provides a wide variety of potential semiconductor/dopant combinations to optimize the ultimate IR detector performance for extrinsic materials.

In general, this preliminary analysis suggests that competitive extrinsic detectors can be found in the III-V family of materials. These might play an important role in high-speed, monolithic-type detector technology, and would probably have significantly better uniformity than HgCdTe in an array application. For ultimate detector performance in very high detectivity applications, cooled extrinsics can have inherent advantages over intrinsic materials; thus, improvements in performance over that of the existing Si extrinsics are also possible. The primary conclusion of this section is that further analysis and investigation of detectors based on extrinsic III-V materials is warranted.

2. Technological Considerations

a) Blocked Impurity Band Detectors

Blocked impurity band (BIB) detectors offer new opportunities for high quantum efficiency in the infrared, especially in the context of an extrinsically doped semiconductor device. Developed at Rockwell Science Center (Petroff and Stapelbroek, 1986; Stetson, *et al.*, 1986) in the early 1980s, counting quantum efficiencies of over 30% were demonstrated in 1987 at wavelengths of 20 microns (Petroff, *et al.*, 1987) (over 50% in the visible) using an Si:As

impurity band detector. In addition, arrays of BIB detectors can reduce electrical crosstalk and increase response uniformity when compared to conventional photoconductor arrays (Stetson, *et al.*, 1986).

Blocked impurity band devices are essentially avalanche photodiodes, with the important difference that they depend on excitations out of the impurity band, instead of the valence band, of a semiconductor. Avalanche multiplication produces internal current gain and allows effective response in photon-starved conditions. Both signal and dark currents, however, are multiplied; thus, operating conditions must be chosen carefully in order to optimize the signal-to-noise ratio and/or equivalent noise power (Sze, 1981).

The detectors are usually fabricated by epitaxial growth. Deposited on top of a degenerately doped semiconductor substrate, which also serves as a contact, is a layer of infrared-active material, followed by a layer of intrinsic (or lightly doped) semiconductor, which is the blocking layer. A shallow ion implant in the blocking layer provides the other (transparent) contact. The total thickness is < 20 microns and, if fabricated out of Si or GaAs, such a device should be naturally radiation-hard.

An external bias causes the infrared-active region to divide into depletion and neutral sections. Incoming infrared photons of appropriate energy (wavelength) excite electron-hole pairs. Pairs created in the depletion region are swept out by the electric field while those in the neutral region recombine. The gain arises from electrons accelerating through the depletion region with sufficient energy to impact-ionize neutral donors (for *n*-type material). The electrons in the conduction band are collected at the transparent contact. The blocking layer permits the external bias to maintain the requisite large internal electric fields in the device (Szmulowicz and Madersz, 1987).

As indicated by the general nature of our discussion of the operating principles of BIB detectors, these devices should be able to supplement and improve the performance of a wide variety of extrinsic semiconductor infrared detectors/arrays. They provide the promise of integrated amplifiers in a monolithic extrinsic Si or GaAs device and they are completely

compatible with the demands of epitaxial fabrication procedures. Given the appropriate band-gap tailoring of the host extrinsic semiconductor, incorporation of blocked impurity-band detector technology may make extrinsic devices preferable to the current intrinsic infrared photodetectors in terms of flexibility and/or performance. The use of extrinsic III-V's, discussed above, in a BIB structure appears to be an ideal match of a material to a device concept, and has apparently not been explored at this time.

b) Refrigeration

Infrared device performance characteristics depend strongly on operating temperature. Hence, improved refrigeration capability is an alternative to advances in materials and device processing in the continuing quest for more responsive and versatile infrared sensors and arrays. For example, both Si:Ga at a temperature of 20K or Si:As at 10K out-perform similarly cooled $\text{Hg}_{1-x}\text{C}_x\text{Te}$, and they allow infrared photon detection in the 3-5, 8-12, and even 12-24 micron wavelength regions. Moreover, well-developed silicon growth technology would permit the immediate fabrication of multi-element, monolithic focal-plane arrays.

Unfortunately, compact and robust refrigerators operating near liquid helium temperatures are not likely to be available soon. Suitable refrigerators operating at or below liquid nitrogen temperatures, however, are a distinct possibility. Decreasing the operating temperature by five or ten Kelvin could make the difference in a comparison between new extrinsic III-V semiconductor devices, strained-layer superlattices or quantum well devices, etc., and MCT, and perhaps allow an alternative to the severe materials constraints imposed by the present generation of infrared detectors.

A complete review of refrigeration technologies is beyond the scope of this document. Rather, we shall attempt to outline some key concerns and promising technologies for refrigerators in the fraction of a watt to one watt cooling-power range appropriate for infrared sensors.

i. Gas-cycle cooling systems

Refrigerators must be reliable on the time-scale of years and, particularly for field operations, reasonably compact. The main breakdown modes are either catastrophic mechanical failure or continuous degradation of operating temperature (which can be $\geq 10\text{K}$ per year). In both cases, there is a clear advantage to limiting the number of moving parts and critical connections. Towards this end, the split Stirling cryocooler has no valves and only a single connecting line between the compressor and the expander. Split-cycle coolers separate the function of the compressor and the cold end for reasons of system packaging and vibration isolation. The Stirling refrigerator's efficiency, however, is sensitive to the length of the connecting line. The Gifford-McMahon cryocooler can have a widely separated compressor assembly and expander assembly (the cold end), but at the cost of two connecting lines plus valves. A third example is the Split Vuilleumier cooler, which is essentially a Stirling refrigerator, except that the mechanical compressor has been replaced by a thermal compressor.

ii. Magnetic cooling systems

Magnetic refrigerators are thermodynamically analogous to gas-cycle refrigerators, where magnetization and magnetic field replace pressure and volume in the Carnot cycle. Good candidate materials should have a large magnetic-spin entropy density, large thermal conductivity, and a reasonably high Debye temperature, so that the spin entropy dominates the entropy of the lattice at low temperatures. Adiabatic demagnetization of copper and praseodymium nickel (PrNi_5) has been used in this manner for decades at very low temperatures ($T < 0.01\text{K}$) and, more recently, single crystals of gadolinium gallium garnet have been used in magnetic cooling cycles below 20K . Reliability should be extremely good, as there are no moving parts whatsoever. There is, however, a residual magnetic field required at the end of the cycle to maintain cooling power. Hence, the infrared detector must either be able to operate in a field or be connected at a distance to the magnetic cooler via a good thermal link.

Recent advances in superconducting materials make this an opportune time for new investment in refrigeration technology. Historically, a maximum superconducting transition

temperature of $T_c = 23\text{K}$ drove refrigeration concerns to $T \leq 10\text{K}$. With the advent of high-temperature superconductivity, $T_c/2 \approx 60\text{K}$, which is a perfect match for the refrigeration requirements of extrinsic semiconductor infrared sensors. The combined interests of the infrared and superconductor device communities should provide new impetus for the development of more efficient and reliable cryocoolers.

III. CONCLUSIONS

In the past five years several potential alternatives to MCT as a material for long-wavelength IR detectors have been introduced. However, the level of development of these systems is far behind that of MCT, so to become competitive more research is required. At present, the two most promising options appear to be strained layer superlattices and extrinsically doped GaAs. The primary reason for pursuing such systems is that III-V technology is significantly more advanced than II-VI technology and the materials are considerably more robust, which makes processing easier and potentially more economic.

Extrinsic III-V semiconductors appear to be ideally suited to the Blocked-Impurity Band detector, and probably should be an area of active interest. However, the utility of extrinsic materials for detectors may ultimately be linked to advances in refrigeration, since in general extrinsic devices must operate at a lower temperature than an intrinsic device to obtain the same performance.

IV. BIBLIOGRAPHY

- D. Ahn and S.-L. Chuang, "Nonlinear intersubband optical absorption in a semiconductor quantum well," *J. Appl. Phys.* **62** (1987) 3052.
- D. Ahn and S.-L. Chuang, "Calculation of linear and nonlinear intersubband optical absorptions in a quantum well model with an applied electric field," *IEEE J. Quantum Elect.* **QE-23** (1987) 2196.
- V. V. Antonov, A. V. Voitsekhovskii, M. A. Frivov, E. V. Malisova, E. N. Mel'chenko, M. P. Nikiforova, E. A. Popova, G. M. Fuks, S. S. Khludkov, "Investigation of the electrical and photoelectric properties of manganese-doped gallium arsenide as a material for photoresistive detectors," *Izvestiya Vysshikh Uchebnykh Zavedenii, Fizika*, **3** (1981) 15 (Sov. Phys. J. **24** (1981) 211).
- F. Bantien, J. Weber, *Sol. State Commun.* **61** (1987) 423.
- Blakemore, J. S., *J. Appl. Phys.* **53** (1982) 520.
- M. M. Blouke, C. B. Burgett, R. L. Williams, *Infrared Phys.* **13** (1973) 61.
- S. E. Botts, "New horizons for IR focal-plane arrays," *Photonics Spectra* **xx** (1988) 125.
- S. J. Brown, Jr., J. S. Blakemore, *J. Appl. Phys.* **43** (1972) 2242.
- R. A. Chapman, W. G. Hutchinson, *Phys. Rev. Lett.* **18** (1967) 443.
- K.-K. Choi, B. F. Levine, C. G. Bethea, J. Walker, and R. J. Malik, "Multiple quantum well $10\text{ }\mu\text{m}$ GaAs/Al_xGa_{1-x}As infrared detector with improved responsivity," *Appl. Phys. Lett.* **50** (1987) 1814.
- W. M. Coderre and J. C. Woolley, "Electrical properties of InAs_{1-x}Sb_x alloys," *Can. J. Phys.* **46** (1968) 1207.
- A. Harwit and J. S. Harris, Jr., "Observation of Stark shifts in quantum well intersubband transitions," *Appl. Phys. Lett.* **50** (1987) 685.
- J. P. Hirth and A. G. Evans, "Damage of coherent multilayer structures by injection of dislocations or cracks," *J. Appl. Phys.* **60** (1986) 2372.
- M. Ilegems, R. Dingle, L.W. Rupp, Jr., *J. Appl. Phys.* **46** (1975) 3059.

- A. M. Jean-Louis and C. Hamon, "Propriétés des alliages $\text{InSb}_{1-x}\text{Bi}_x$," *Phys. Stat. Sol.* **34** (1969) 329.
- M. A. Kinch, "Materials requirements for IR detectors and imagers," *Mats. Res. Soc. Symp. Proc.* **90** (1987) 15.
- S. R. Kurtz, G. C. Osbourn, R. M. Biefield, L. R. Dawson, H.J. Stein, "Extended infrared response of InAsSb strained-layer superlattices," *Appl. Phys. Lett.* **52** (1988) 831.
- D. V. Lang, R. A. Logan, *J. Electron. Mater.* **4** (1975) 1053.
- B. F. Levine, R. J. Malik, J. Walker, K.-K. Choi, C. G. Bethea, D. A. Kleinman, and J. M. Vandenberg, "Strong $8.2\ \mu\text{m}$ infrared intersubband absorption in doped GaAs/AlAs quantum well waveguides," *Appl. Phys. Lett.* **50** (1987) 273.
- B. F. Levine, K.-K. Choi, C. G. Bethea, and R. J. Malik, "New $10\ \mu\text{m}$ infrared detector using intersubband absorption in resonant tunneling GaAlAs superlattices," *Appl. Phys. Lett.* **50** (1987) 1092.
- B. F. Levine, K.-K. Choi, C. G. Bethea, J. Walker, and R. J. Malik, "Quantum well avalanche multiplication initiated by $10\ \mu\text{m}$ intersubband absorption and photoexcited tunneling," *Appl. Phys. Lett.* **51** (1987) 934.
- D. Long, *Infrared Phys.* **7** (1967) 121.
- D. Long, in *Topics in Applied Physics: Optical and Infrared Detectors*, Vol. 19, Springer-Verlag, Berlin, 1980.
- F. J. Low, *J. Opt. Soc. Am.* **51** (1961) 1300.
- F. J. Low, T. Nishimara, A. W. Davidson, and M. Alwardi, in *Proc. Infrared Astronomy with Arrays: Workshop on Ground-Based Astronomical Observations with Infrared Detectors, University of Hawaii at Hilo, 24-26 March, 1987*, Eds. C. G. Wynn-Williams and E. E. Becklin.
- D. H. Martin and D. Bloor, *Cryogenics* **1** (1961) 159.
- A. Mitonneau, G.M. Martin, A. Mircea, *Electron. Lett.* **13** (1977) 667.

- L. Montelius, S. Nilsson, L. Samuelson, E. Janzen, J. Ahlstrom, Appl. Phys. **64** (1988) 1564.
- A. J. Noreika, W. J. Takei, M. H. Francombe, and C. E. C. Wood, "Indium antimonide-bismuth compositions grown by molecular beam epitaxy," J. Appl. Phys. **53** (1982) 4932.
- G. C. Osbourn, "InAsSb strained-layer superlattices for long wavelength detector applications," J. Vac. Sci. Technol. B **2** (1984) 176.
- G. C. Osbourn, "Recent trends in III-V strained layer research," J. Vac. Sci. Technol. B **4** (1986a) 176.
- G. C. Osbourn, "Strained-layer superlattices: A brief review," IEEE J. Quantum Elec. QE-22 (1986b) 1677.
- M. D. Petroff and M. G. Stapelbroek, *Blocked Impurity-Band Detectors*, US Patent No. 4568960, filed 10/23/80, granted 2/4/86.
- M. D. Petroff, M. G. Stapelbroek and W. A. Kleinhaus, Appl. Phys. Lett. **51** (1987) 406.
- M. Y. Pines, R. Baron, in *Proc. IEEE Int. Elect. Dev. Meeting*, 1974, p.446.
- Y. Sasaki, T. Sato, K. Matsushita, T. Hariu, Y. J. Shibata, Appl. Phys., **57** (1985) 1109.
- N. Sclar, Infrared Phys. **16** (1976) 435.
- R. A. Soref, J. Appl. Phys. **38** (1967) 5201.
- S. B. Stetson, D. B. Reynolds, M. G. Stapelbroek, and R. L. Stormer, Proc. SPIE **686** (1986) 48.
- S. M. Sze, *Physics of Semiconductor Devices*, 2nd Ed., (J. Wiley and Sons, New York: 1981).
- F. Szmulowicz and F. L. Madarsz, J. Appl. Phys. **62**, 2533 (1987).
- L. C. West and S. J. Eglash, "First observation of an extremely large-dipole infrared transition within the conduction band for a GaAs quantum well," Appl. Phys. Lett. **46** (1985) 1156.
- D. A. Woodbury, J. S. Blakemore, Phys. Rev. **B 8** (1973) 3803.
- K. M. van Vliet, Appl. Opt. **6** (1967) 1145.

Appendix: Details of Calculations for D_{λ}^* Estimates for Mn/GaAs

A. Thickness

Calculate absorption coefficient first:

Use Equation (6) with $\sigma = 10^{-16} \text{ cm}^2$ (Chapman and Hutchinson, 1967) and

$$(N_A - N_D) = 2.5 \times 10^{18} \text{ cm}^{-3}$$

Yields: $\alpha = 2.5 \times 10^2 \text{ cm}^{-1}$

Use: $t = 2.5 \alpha^{-1}$ for 90% absorption

Yields: $t = 50 \mu\text{m}$ (with a back reflector for two optical passes through the active region)

B. Calculate the value of p_1

Use: $N_v = 1.12 \times 10^{18} \text{ cm}^{-3}$ (Blakemore, 1982) and $E_t - E_v = 0.11 \text{ eV}$; $T = 77 \text{ K}$

Yields: $p_1 = 7.2 \times 10^{10} \text{ cm}^{-3}$

C. Calculate value of σ_p

Use: $\sigma_p = 3.8 \times 10^{-15} \text{ cm}^2$ at 45.7 K and a $T^{-3.9}$ temperature dependence (Montelius, *et al.*, 1988)

Yields: $\sigma_p = 4.6 \times 10^{-16} \text{ cm}^2$ at 77 K

D. Calculate B

Use: v_{th} approx $7 \times 10^6 \text{ cm/sec}$ (Montelius, *et al.*, 1988) and $\sigma_p = 4.6 \times 10^{-16} \text{ cm}^2$

Yields: $B = 3.12 \times 10^{-9} \text{ cm}^3/\text{sec}$

E. Calculate $p_1 B (N_A - N_D) t$

Use above values in parts A-D,

Yields: $2.8 \times 10^{18} \text{ cm}^{-2} \text{ s}^{-1} = 2.8 \times 10^{22} \text{ m}^{-2} \text{ s}^{-1}$

F. Calculate D_{λ}^* assuming detector limit

Use: $2.8 \times 10^{22} \text{ m}^{-2} \text{ s}^{-1}$ as value of [] term in Equation (1)

Yields: $D_{\lambda}^* = 1.9 \times 10^{10} \text{ cm-Hz}^{1/2}/\text{W}$

Note: Because the product $(N_A - N_D)t$ appears in the expression for B, the D_{λ}^* value in section F does not depend on achieving $N_A - N_D = 2.5 \times 10^{18} \text{ cm}^{-3}$, provided a thicker detector element can be tolerated.

G. Calculate D_{λ}^* at different temperatures (50 K, 20 K)

New values for p_1 and σ_p (at different temperatures) were obtained without taking into account any temperature dependences in N_v or v_{th} . This should provide rough estimates which will require experimental confirmation before they are accepted.

TABLE 1. System Comparison Estimates

For data taken at 77K, 2π FOV:

$$\begin{aligned}(\text{BLIP}) D_{\lambda}^* &= 4.51 \times 10^{10} \text{ cm-Hz}^{1/2}/\text{W} \\ \text{Mn:GaAs } D_{\lambda}^* &= 1.87 \times 10^{10} \text{ cm-Hz}^{1/2}/\text{W} \\ \text{HgCdTe } D_{\lambda}^* &= 4.85 \times 10^{10} \text{ cm-Hz}^{1/2}/\text{W} \\ \text{Extrinsic Si } D_{\lambda}^* &<< 10^9 \text{ cm-Hz}^{1/2}/\text{W}\end{aligned}$$

For data taken at 50K, 2π FOV:

$$\begin{aligned}(\text{BLIP}) D_{\lambda}^* &= 4.5 \times 10^{10} \text{ cm-Hz}^{1/2}/\text{W} \\ \text{Mn:GaAs } D_{\lambda}^* &= 9.41 \times 10^{11} \text{ cm-Hz}^{1/2}/\text{W} \\ \text{HgCdTe } D_{\lambda}^* &= 1 \times 10^{12} \text{ cm-Hz}^{1/2}/\text{W} \\ \text{Extrinsic Si } D_{\lambda}^* &< 10^9 \text{ cm-Hz}^{1/2}/\text{W}\end{aligned}$$

For data taken at 20K, 2π FOV:

$$\begin{aligned}(\text{BLIP}) D_{\lambda}^* &= 4.5 \times 10^{10} \text{ cm-Hz}^{1/2}/\text{W} \\ \text{Mn:GaAs } D_{\lambda}^* &= > 10^{13} \text{ cm-Hz}^{1/2}/\text{W} \\ \text{HgCdTe } D_{\lambda}^* &= 1 \times 10^{12} \text{ cm-Hz}^{1/2}/\text{W} \\ \text{Extrinsic Si } D_{\lambda}^* &< 5 \times 10^{13} \text{ cm-Hz}^{1/2}/\text{W}\end{aligned}$$

Table 2 - Properties of Important Semiconductors (from Sze, 1981)

Semiconductor		Bandgap (eV)		Mobility at 300 K (cm ² /V-s) ^a		Band ^b	Effective Mass <i>m</i> [*] / <i>m</i> ₀		ϵ/ϵ_0
		300 K	0 K	Elec.	Holes		Elec.	Holes	
Element	C	5.47	5.48	1800	1200	I	0.2	0.25	5.7
	Ge	0.66	0.74	3900	1900	I	1.64 ^c 0.082 ^d	0.04 ^e 0.28 ^f	16.0
	Si	1.12	1.17	1500	450	I	0.98 ^c 0.19 ^d	0.16 ^e 0.49 ^f	11.9
	Sn		0.082	1400	1200	D			
IV-IV	α -SiC	2.996	3.03	400	50	I	0.60	1.00	10.0
III-V	AlSb	1.58	1.68	200	420	I	0.12	0.98	14.4
	BN	-7.5				I			7.1
	BP	2.0							
	GaN	3.36	3.50	380			0.19	0.60	12.2
	GaSb	0.72	0.91	5000	850	D	0.042	0.40	15.7
	GaAs	1.42	1.52	8500	400	D	0.067	0.082	13.1
	GaP	2.26	2.34	110	75	I	0.82	0.60	11.1
	InSb	0.17	0.23	80000	1250	D	0.0145	0.40	17.7
	InAs	0.36	0.42	33000	460	D	0.023	0.40	14.6
	InP	1.35	1.42	4600	150	D	0.077	0.64	12.4
II-VI	CdS	2.42	2.56	340	50	D	0.21	0.80	5.4
	CdSe	1.70	1.85	800		D	0.13	0.45	10.0
	CdTe	1.56		1050	100	D			10.2
	ZnO	3.35	3.42	200	180	D	0.27		9.0
	ZnS	3.68	3.84	165	5	D	0.40		5.2
IV-VI	PbS	0.41	0.296	600	700	I	0.25	0.25	17.0
	PbTe	0.31	0.19	6000	4000	I	0.17	0.20	30.0

^a The values are for drift mobilities obtained in the purest and most perfect materials available to date.

^b I = indirect, D = direct.

^c Longitudinal effective mass.

^d Transverse effective mass.

^e Light-hole effective mass.

^f Heavy-hole effective mass.

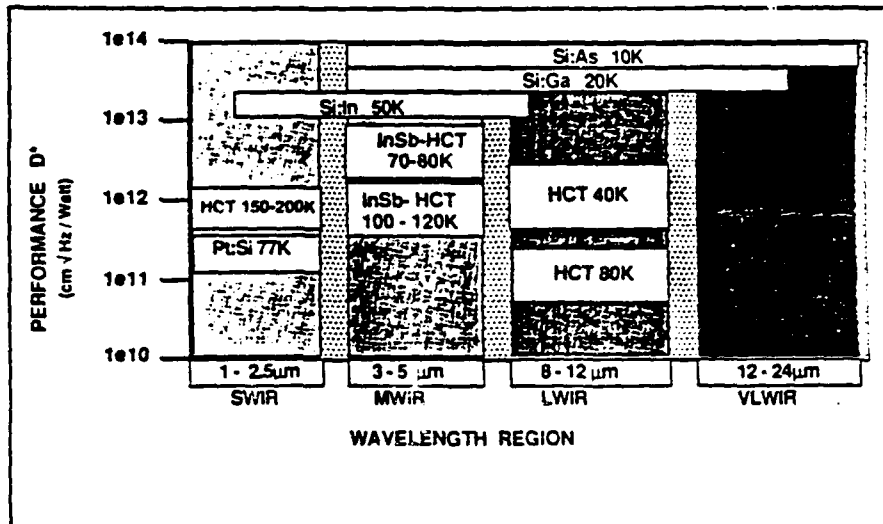
Table 3 - Lattice Constants

	Element or Compound	Name	Crystal ^a Structure	Lattice Constant at 300 K (Å)
Element	C	Carbon (diamond)	D	3.56683
	Ge	Germanium	D	5.64613
	Si	Silicon	D	5.43095
	Sn	Grey Tin	D	6.48920
IV-IV	SiC	Silicon carbide	W	$a = 3.086, c = 15.117$
III-V	AlAs	Aluminum arsenide	Z	5.6605
	AlP	Aluminum phosphide	Z	5.4510
	AlSb	Aluminum antimonide	Z	6.1355
	BN	Boron nitride	Z	3.6150
	BP	Boron phosphide	Z	4.5380
	GaAs	Gallium arsenide	Z	5.6533
	GaN	Gallium nitride	W	$a = 3.189, c = 5.185$
	GaP	Gallium phosphide	Z	5.4512
	GaSb	Gallium antimonide	Z	6.0959
	InAs	Indium arsenide	Z	6.0584
	InP	Indium phosphide	Z	5.8686
	InSb	Indium antimonide	Z	6.4794
II-VI	CdS	Cadmium sulfide	Z	5.8320
	CdS	Cadmium sulfide	W	$a = 4.16, c = 6.756$
	CdSe	Cadmium selenide	Z	6.050
	CdTe	Cadmium telluride	Z	6.482
	ZnO	Zinc oxide	R	4.580
	ZnS	Zinc sulfide	Z	5.420
	ZnS	Zinc sulfide	W	$a = 3.82, c = 6.26$
IV-VI	PbS	Lead sulfide	R	5.9362
	PbTe	Lead telluride	R	6.4620

^aD = Diamond, W = Wurtzite, Z = Zincblende, R = Rock salt.

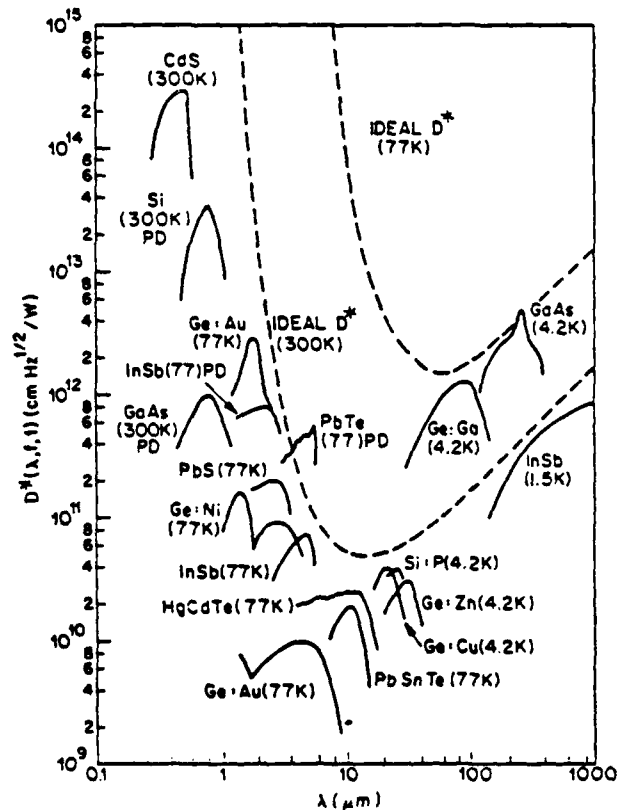
FIGURE 1 - Performance of Popular Modern Detectors

a)



Operating regions of popular modern detectors (from Botts, 1988).

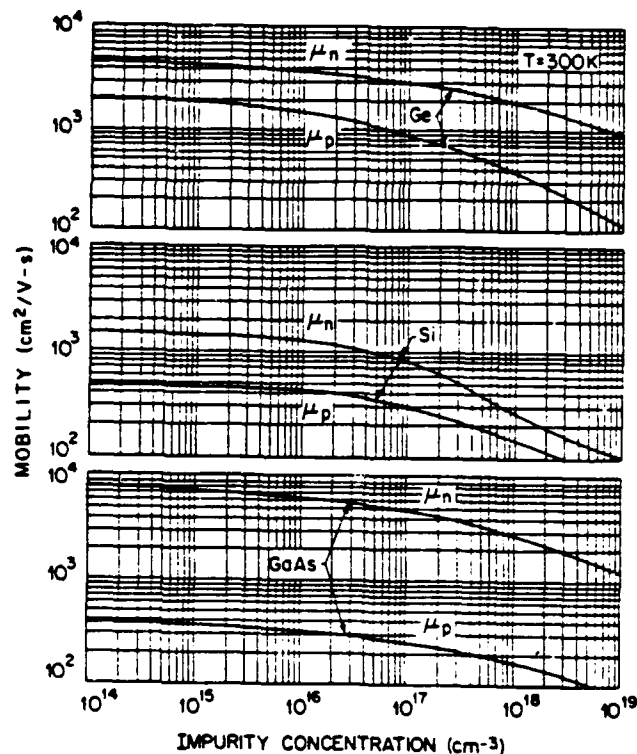
b)



Detectivity D^* as a function of wavelength for various photoconductors and photodiodes (indicated with PD). The dashed curves are the theoretical ideal D^* at 77 L and 300 K viewing an angle of 2π steradians (from Sze, 1981).

FIGURE 2

a) Drift mobility of Ge, Si and GaAs at 300 K versus impurity concentration. (After Casey and Panish; Prince; Beadle, Plummer and Tsai.)



b)

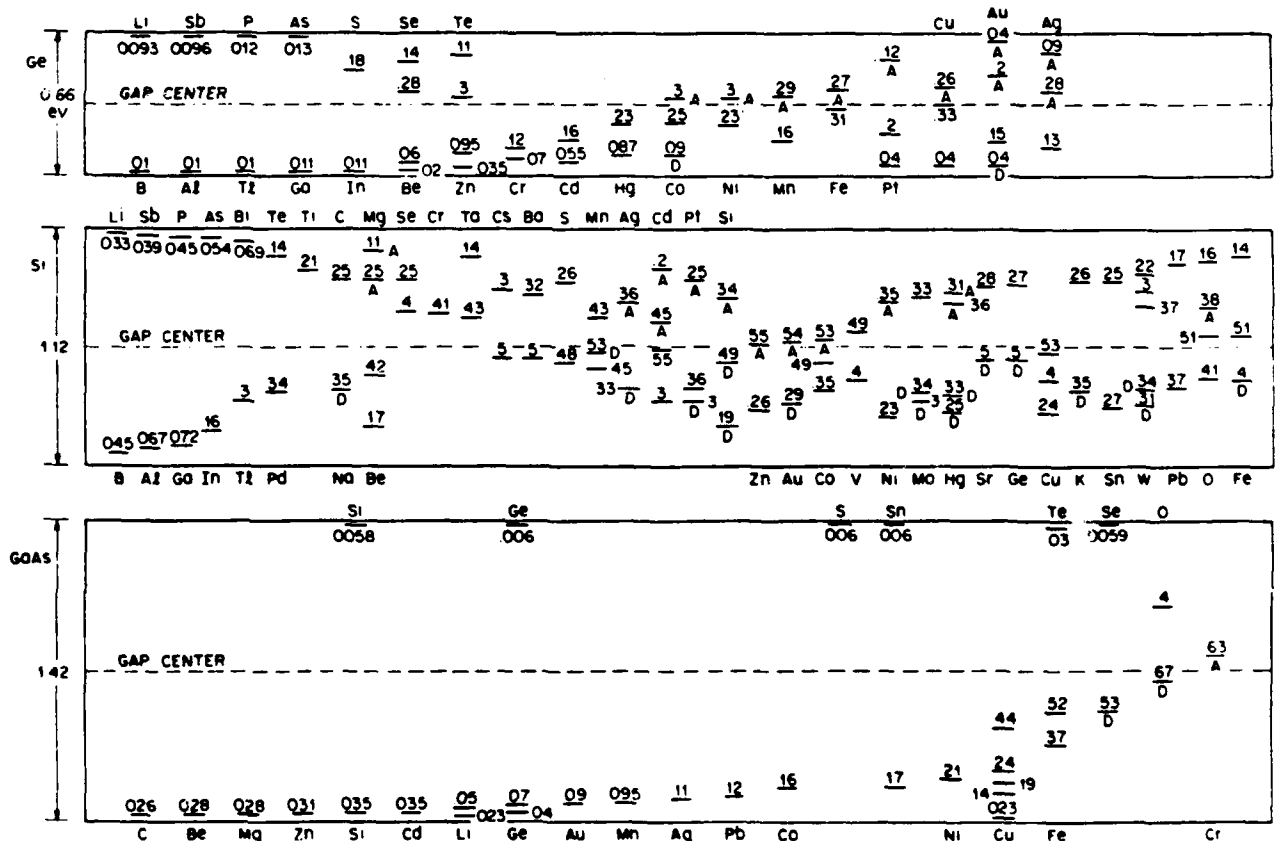
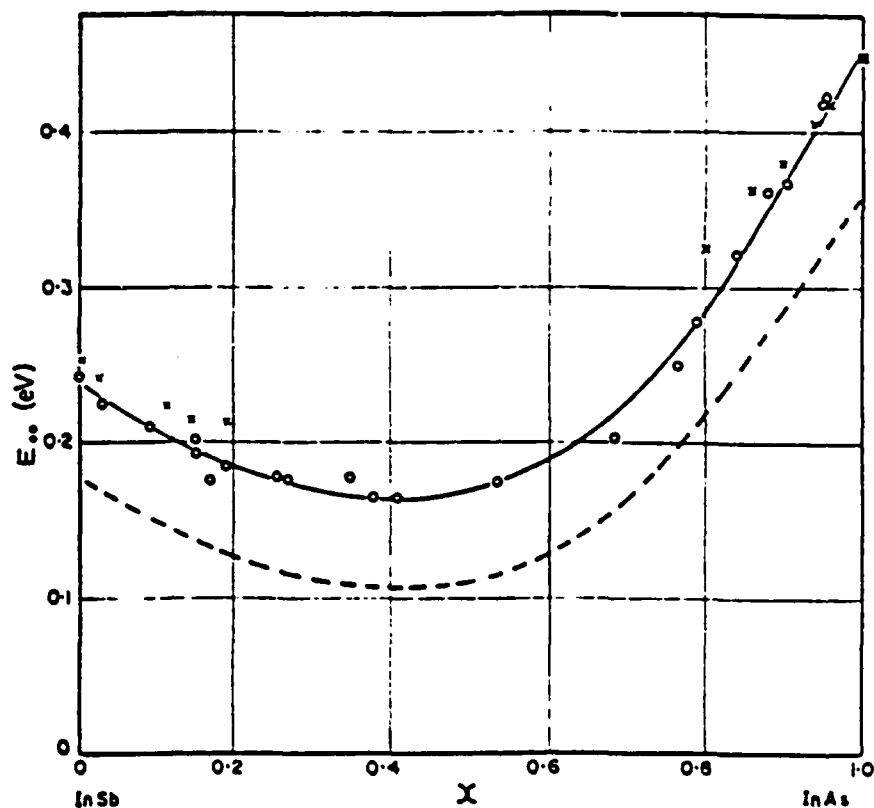


Fig. 13 Measured ionization energies for various impurities in Ge, Si, and GaAs. The levels below the gap centers are measured from the top of the valence band and are acceptor levels unless indicated by D for donor level. The levels above the gap centers are measured from the bottom of the conduction-band level and are donor levels unless indicated by A for acceptor level. The bandgaps at 300 K are 0.66, 1.12, and 1.42 eV for Ge, Si, and GaAs, respectively. (After Conwell, Ref. 27; Sze and Irvin, Ref. 28; Milnes, Ref. 24.)

(from Sze, 1981).

FIGURE 3

- a) Variation of extrapolated absolute-zero band gap E_{00} with composition x .
 O, present results; \times , extrapolated absolute-zero band gap E_{g0} from optical data (Woolley and Warner 1964);
 ---, room-temperature optical band gap E_g (Woolley and Warner, 1964) - from Coderre and Wiley, 1968.



- b) Variation of the forbidden band E_G^0 with composition x , where x = mole fraction InBi in InSb. (Jean-Louis and Hamon, 1969)

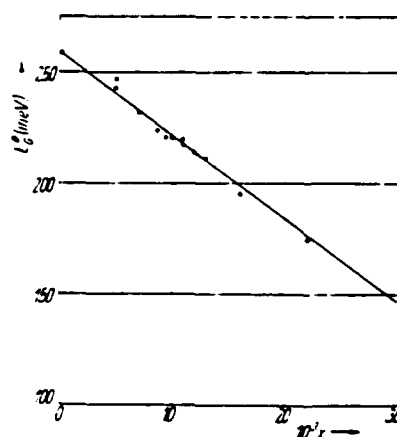
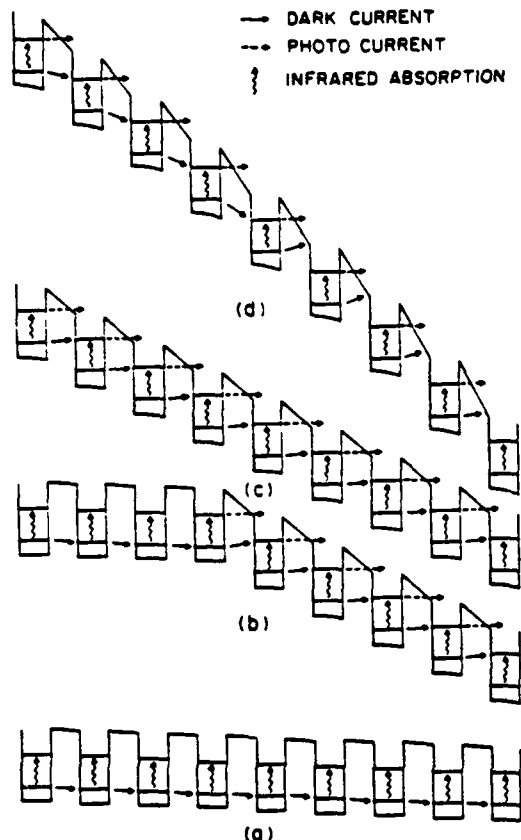


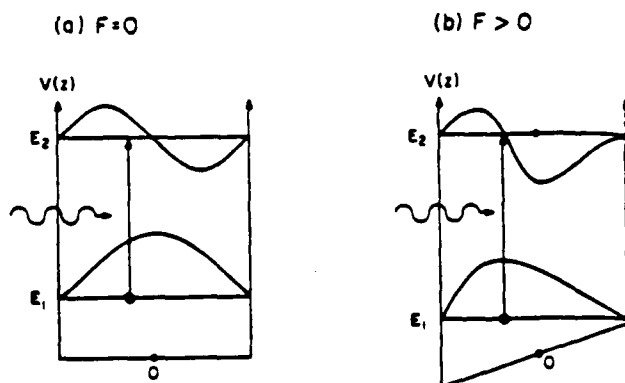
FIGURE 4

— DARK CURRENT
 --- PHOTO CURRENT
 ? INFRARED ABSORPTION

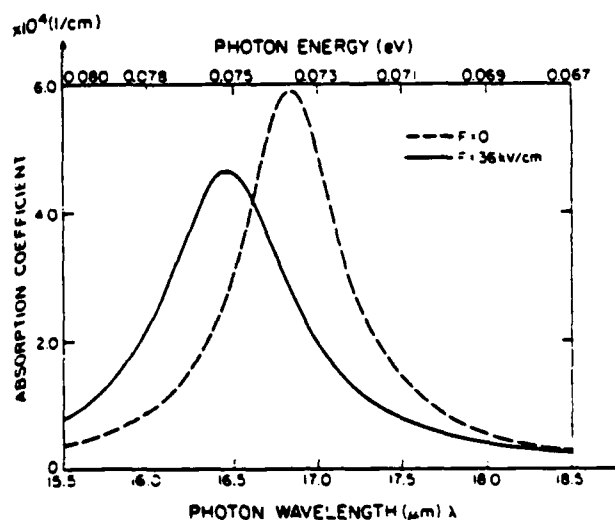


a) Potential profiles of the device at bias V_g : (a) less than 0.3 V, (b) between 0.3 and 3.4 V; (c) between 3.4 and 5.8 V, and (d) between 5.8 and 9 V.

(K.-K. Choi, *et al.*, 1987)



b) Schematic diagrams for energy levels and wave functions of an infinite quantum well for (a) $F=0$ and (b) $F>0$.



c) The linear intersubband absorption coefficient α ($I=0$) for an infinite well with an effective width $L=151.2 \text{ \AA}$ with $5.83 \times 10^{17} / \text{cm}^2$ electrons at $T=90 \text{ K}$, with a field-dependent linewidth plotted for the zero electric field case - - - and an applied electric field of 36 kV/cm —. Theoretical results are in good agreement with experimental results.

(Ahn and Chuang, 1987)

FIGURE 5 - Schematic view of detector cross section (not to scale!)

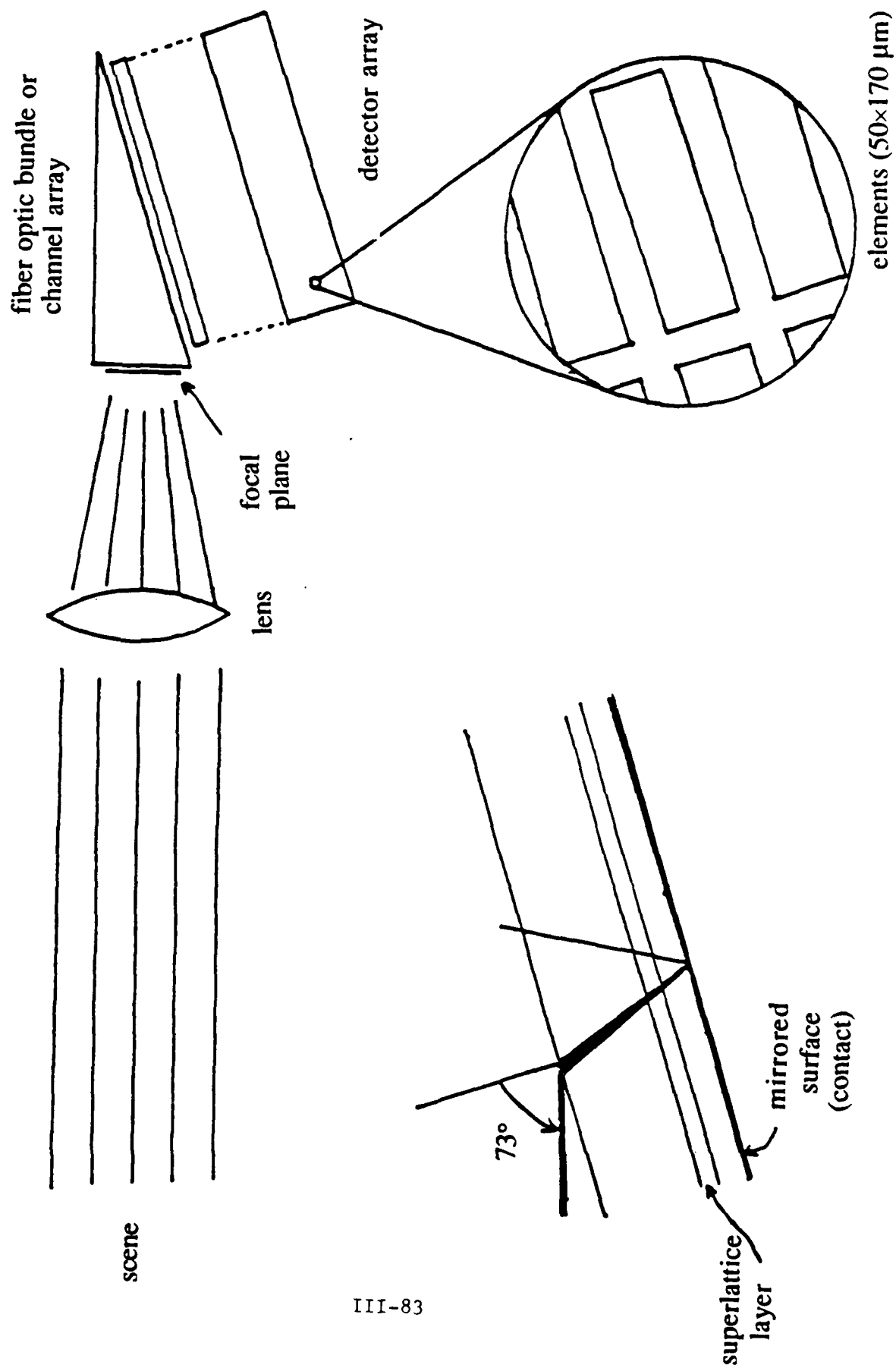


FIGURE 6 - Schematic of an SLS (from Osbourne, 1986a)

Strained-Layer Superlattice

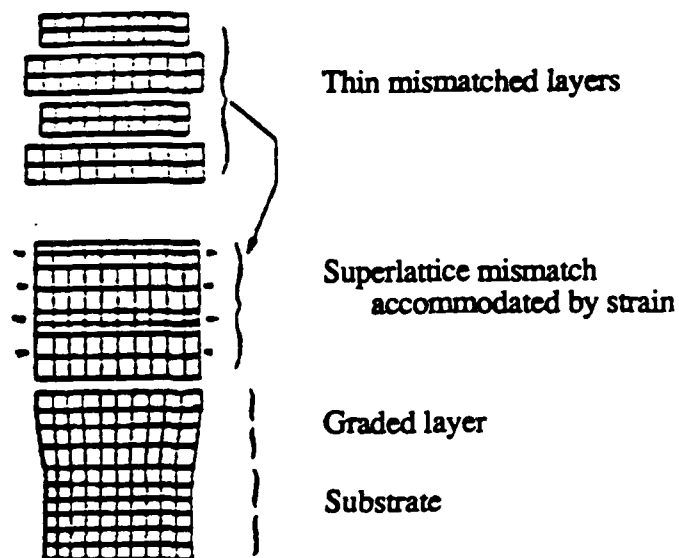
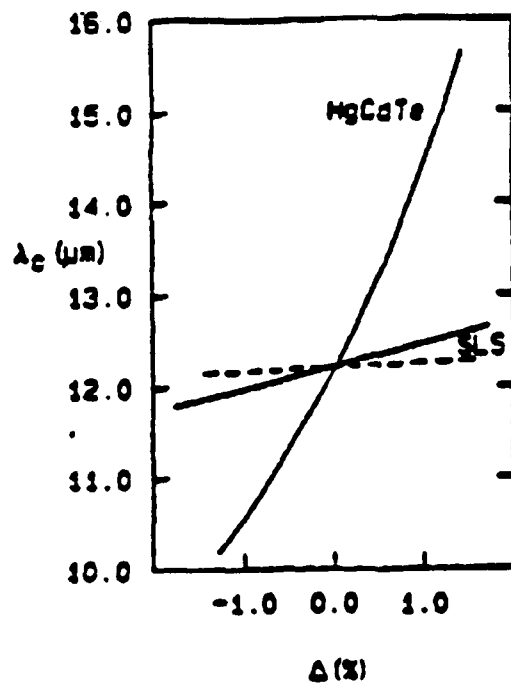


FIGURE 7 - after Osbourne (1984)



Calculated λ_c values at 77 K for $\text{InAs}_{0.39}\text{Sb}_{0.61}/\text{InAs}_{1-y-\Delta}\text{Sb}_{y+\Delta}$ SLS's (—) and $\text{InAs}_{0.39-\Delta}\text{Sb}_{0.61+\Delta}/\text{InAs}_{1-y-\Delta}\text{Sb}_{y+\Delta}$ SLS's (- - - -) with 500Å layers, which are equally thick with $y=0.82$ as a function of Δ . For comparison, the λ_c variation of bulk $\text{Hg}_{0.795-\Delta}\text{Cd}_{0.204+\Delta}\text{Te}$ at 77K is also include.

C. USES OF CHARGED PARTICLE BEAMS FOR TRANSPORTABLE HF RADAR

Warren S. Warren

I. Introduction

Radar systems must be continually upgraded to simply maintain our current defensive posture, since potential threats such as low-flying aircraft or cruise missiles have ever more sophisticated ECM capabilities and reduced cross-sections. Much of the recent developmental work has been concentrated at ever higher microwave frequencies, in part because of recent design improvements for components operating near 100 GHz, and in part because a higher carrier frequency permits larger absolute bandwidths (useful for high repetition rate and robustness to jamming) for the same fractional bandwidth (which is a critical parameter in antenna and detector design).

However, the opposite end of the microwave spectrum also has its advantages:

- Using multiple radar frequencies significantly complicates cross-section reduction, particularly if the frequencies are vastly different.

- Radar systems operating at frequencies below roughly 30 MHz are partially reflected by the ionosphere, so the normal limitations of ground-based systems imposed by the earth's curvature can be eliminated. For example, the Over-the-Horizon Backscatter (OTH-B) radar system works at HF frequencies to take advantage of ionospheric reflections, and provides long-range, all-altitude tactical early warning and surveillance.

- Any possible desired manipulation of radiofrequency (<1 GHz) waveforms is simple, highly accurate, and cheap. Completely arbitrary waveform generation (any phase and amplitude modulation) is straightforward, and amplification of even a wide bandwidth waveform to the several watt level basically requires a single transistor. Amplitude adjustments with at least 80 dB of dynamic range and phase shift increments of 0.1° are readily done. Components can be interconnected by coaxial cable with minimum loss. Filters can be built with fractional bandwidths of 0.01% on up using only capacitors and inductors, or narrower using crystals. All of the components necessary for these manipulations are available at costs no greater than \$500, and generally at least an order of magnitude less.

Unfortunately, efficient transmission or detection of HF waves in air is somewhat complicated, since wavelengths are relatively long ($\lambda=30$ m for $\nu=10$ MHz). The transmit and

receive antennas for one OTH-B system are shown in Figures 1 and 2. The broadband nature accounts for some of the size of these antennas; the carrier frequency is adjusted to the maximum frequency which still gives significant return, and this is a strong function of ionospheric conditions. In addition, requirements for high power transmission and high reception sensitivity requirements of these antennas force extended geometries. However, they are obviously not very portable. In addition, restrictions to small fractional bandwidth signals and physical limitations imposed by the need to use Doppler information make the waveforms quite predictable.

The overwhelming predominance of radars using small fractional bandwidth signals dates back at least fifty years, and was originated by the elementary fact that passive linear components (capacitors, inductors and resistors) show resonances that are exponentially damped sine or cosine waves. Such resonant circuits reject background noise at other frequencies and tremendously enhance detection capabilities. However, resonant circuits for nonsinusoidal waveforms can be built with active components, such as operational amplifiers. For example, a circuit could be built which transmits only the fourth Legendre polynomial, and attenuates all others. This has the practical advantage that the relative bandwidth is not limited by the properties of resonant circuits to be a very small number; enlarged bandwidths increase the rate of information transmission and make duplication more difficult.

If wide bandwidth radar (WBR) transmission in the HF region is acceptable, a novel and potentially useful transmitter would be a charged particle beam (CPB). As discussed in section II, a beam of relativistically fast electrons produces a highly directional electromagnetic pulse with enormous energy. Depending on initial conditions, this pulse will typically last a few tens of nanoseconds, thus giving frequency components up to a few tens of megahertz ($\Delta\nu \Delta t \approx 1$). A CPB would be much more portable and concealable than the massive fixed antennas described above.

This report discusses the advantages and limitations of WBR transmitters in general and CPBs in particular (Section III), explores the alternatives for detecting the return signal (Section IV), and points out the restrictions this approach imposes on radar countermeasure design (Section V). We conclude that CPBs with reasonably attainable performance characteristics present an extremely attractive and robust alternative for HF generation in radar applications.

II. Charged Particle Beams as HF generators

The CPB is essentially a long (several meters), thin (several millimeters), cylindrical packet of charge which has been accelerated to relativistic velocities ($\beta=v/c \approx 1$) and propagates through

space. Figure 3 shows a typical beam waveform. Extremely high peak currents (≈ 10 kA) can be achieved. In the absence of a target, the propagation distance is limited by the initial energy and by the onset of instabilities which ultimately destroy its coherence; other phenomena reduce the effective length and speed of the charge distribution as it propagates.

The fields generated by this charge distribution can be conveniently decomposed into several contributions. Let us start with a simple geometry (Figure 4) including no ground planes or reflections. The simplest field contribution, the "start pulse", comes from the appearance of charge at the accelerator exit port. The field close to (and radiated away from) the exit port is polarized in the θ direction, and is given by

$$E_{\theta}(r, \theta; t) = (Z_0 \beta \sin \theta / 4\pi r (1 - \beta \cos \theta)) I_0(t - r/c) \quad [1]$$

where $Z_0 = 120\pi \Omega$ is the impedance of free space. This field distribution is highly anisotropic (it is a Lorentz-transformed Coulomb field) and is peaked in the direction $\theta_{\max} = \cos^{-1} \beta$. Figure 5 shows the normalized electric field patterns as a function of β . For $\beta = 0.95$, which is in fact attainable in the atmosphere, Figure 6 shows constant-field contours and indicates the staggering fields involved. For a peak current of 10 kA, the peak field 10 m from the exit port is 90 kV/m. The temporal profile replicates Figure 3. For comparison, the high-altitude nuclear electromagnetic pulse environment specified in Mil-Std-2169, "High-Altitude Electromagnetic Pulse (HEMP) Environment"(U) peaks at over 50 kV/m and rises in a few ns¹. Thus there is a significant EMP threat posed by a CPB.

Of more concern for the purposes of this report is the field at a distance. Equation [1] predicts a field which is still 100 V/m 9 km away. The frequency distribution is given by the Fourier transform of the pulse envelope in Figure 3. The Fourier transform of a rectangle of length T would be proportional to $\sin(\omega T/2)/(\omega T)$; this waveform will produce a similar frequency spectrum, with the highest frequency lobes truncated by the finite rise and fall times. The first frequency zero for this 40 ns pulse appears at $\omega = 50\pi$ MA $v^2 = 25$ MHz, and the side lobes will extend out to about $\pi/(4 \text{ ns}) \approx 800$ MA $v = 130$ MHz. Thus the pulse transmits at all frequencies up to the HF range.

There are other contributions to the field as well. The beam carries a field with it as it propagates (this is not a radiated field; no accelerations are involved). As the beam decelerates one gets a "stop pulse" which is opposite in sign to the start pulse and is calculable if the deceleration is known. Finally, a ground reflected signal of opposite sign will always follow. Typical ground dielectric constants are of order 15 at relevant frequencies. Amplitude reflection coefficients

depend on the angle of incidence, but are generally $\leq 0.7^3$. The reflected pulse will in general be somewhat distorted, and will follow the initial pulse by a time delay $\Delta t = (2h/c) \sin \chi$, where χ is the observer's elevation angle above the horizon. These reflections will change the relative magnitudes of the Fourier components of the pulse, and so the spectrum can be tailored to some extent.

III. Wide Bandwidth Radar Theory

A. Applications of nonsinusoidal waveforms

A small community of scientists has been working for years on applications of nonsinusoidal waveforms to radar and communications. This work has been quite controversial, with heated arguments on both sides of the issue; these arguments have tended to invoke specious logic and qualitative generalizations rather than defensible calculations. Rather than recount the history of this field, we note that a recent report by scientists at the Environmental Research Institute of Michigan⁴ seems to have done an excellent job of sorting between competing claims, and we review a few of their conclusions here.

There are no fundamental obstacles to the use of nonsinusoidal waveforms in radar systems. Component design is more complicated than for narrow bandwidth systems, so existing devices do not have the same performance as do (conventional) narrowband amplifiers and detectors; on the other hand, the amount of developmental time invested in the conventional devices is many orders of magnitude greater. For example, consider the problem of making an optimum receiver. With a single frequency signal this is straightforward, since narrowband resonant circuits are easily built. However, optimum filter functions can be calculated for any expected frequency distribution of the returned signal. Such filter functions are achieved in a wide variety of applications requiring sensitive detection (such as nuclear magnetic resonance, which works in the 100-600 MHz frequency range) by bounding the signal spectrum with a broadband filter and convoluting the time domain return with the optimum filter function (this requires only fast Fourier transforms).

Reference (6) also sets up a detailed mathematical model of a radar detection problem (an aircraft in the presence of clutter and thermal noise) and calculates the waveform which optimizes the detection probability as a function of pulse energy and other parameters. In a radar application we assume the field intensities are insufficient to cause nonlinear effects in their targets (in other words, at these distances we are trying to detect objects, not explode them), so the reflected signal at frequency ω will be doubled if the pulse Fourier component at frequency ω is doubled.

Nonetheless, the intuitively obvious extrapolation that the optimum pulse shape is independent of power is wrong: in general, the optimum shape has a broader frequency spectrum as the pulse energy increases. This comes from the effects of clutter and thermal noise. At very low intensities it is best to overwhelm the thermal background by concentrating energy in a small range of Fourier components near the frequency of maximum target cross section. At higher energies we can take advantage of the difference between the frequency dependence of target and clutter cross-sections by broadening the frequency spectrum; we no longer need narrowband signals to overwhelm the noise. The exact details of the optimum pulse broadening are a strong function of the target resonances, assumed radar system parameters, and clutter distribution, but the general conclusion is clear: *in the limit of very high pulse energies, optimum target detection is provided by very broadband signals. This is of course the limit of operation of a CPB.*

B. Constraints on the Transmitted HF Waveform to Retrieve Range or Velocity and Optimize Signal/Noise Ratios

Simple considerations about the nature of the target and the frequency distribution impose fairly strong constraints on the nature of the transmitted waveform which can be used for range or velocity determination. The simplest case to consider is a radar which simply transmits short pulses with some repetition rate, such as a system based on a CPB; the arguments can be readily generalized to describe a more conventional frequency swept system. For the sake of illustration we will focus on applications to OTH radar, but a similar analysis can readily be performed for other radar applications.

1. The **maximum repetition rate** to avoid a range ambiguity is set by the transit time between the transmitter antenna, the target, and the receiver antenna. In an OTH application detection has to occur over distances up to roughly 3000 km; it takes the radiowaves 20 ms to travel that distance and back in a straight line (slightly longer if the real path with reflections is considered.) Most other applications can use a higher repetition rate, if it is available.

2. There is also a **minimum repetition rate** for the periodic waveform to distinguish signal from clutter via the Doppler effect. Again as an example, OTH radar distinguishes target from ocean, which of course has a much larger total cross section, by Doppler shift. The Doppler shift at 20 MHz of a cruise missile heading directly at the radar at 1000 km/hr is 40 Hz; any other direction gives a smaller shift. A single CPB pulse (or for that matter a conventional frequency swept radar) covers a bandwidth which is much larger than 40 Hz, so how can cruise missile reflections of the 10.000000 MHz radar component be distinguished from ocean reflections of the 10.000040 MHz component? They can't. However, if the CPB pulse is *exactly* repeated at a rate of X pulses per

second its Fourier transform only contains frequency components at X Hz intervals. The Doppler shifted signal falls between the transmitted frequencies if the repetition rate is sufficiently high, so it can be resolved in principle by signal processing; if the repetition rate is too low there are "blind speeds".

With current technology, CPB repetition rates of tens to hundreds of hertz are attainable. As noted earlier, any higher repetition rate would be wasted for an OTH application, although it might be useful in other radar applications. One advantage of a wideband signal is that blind speed is proportional to frequency. Thus any possible speed will have a set of WBR reflections which differ from the reflections induced by static clutter.

3. The **minimum number of repetitions** to acquire a signal is constrained as well. To see this we can again think in terms of Fourier transforms and OTH applications: the return from the missile might typically be 60-80dB smaller than the ocean return, so terminating the two returns after a time T must not spend much of the frequency distribution associated with the ocean return into the frequency distribution associated with the missile return. Terminating the return is exactly the same as multiplying the time domain waveform by a square pulse of length T, or equivalently in the frequency domain convoluting the infinite-repetition frequency distribution (which consists of a series of delta functions) with the sinc function

$$F(\omega) = 2 (\sin (\omega T/2))/(\omega T)$$

which has maxima of magnitude $2/\omega T$. The ocean return will be 60 dB down ($F(\omega)=.001$) 40 Hz away only if $T > 8$ sec; it will be 80 dB down only if $T > 80$ sec. The ratio (power at the peak frequencies)/(power far from peak frequencies) scales as the square of the number of pulses. A still better scaling would be achieved by a rotating transmitter, which makes the pulse train intensity rise and fall gradually at any position without discontinuities. The beam lobes in Figure 5 would be multiplied by the rotation rate to give the temporal profile.

This argument assumes that the signal is much smaller than background reflections, which is reasonable for OTH but may be an unnecessarily strict limitation for other radar applications. It also assumes the waveform is perfectly reproduced from shot to shot. With current technology, CPB pulses achieve a "flat top" to better than 1% accuracy, and have comparable reproducibility. This level of jitter gives a Fourier spectrum with well defined peaks spaced by the repetition rate, but superimposed on these peaks is an essentially white spectrum. Again to use a concrete example, if the waveform were repeated at 80 Hz and terminated after 8 sec, the frequency spectrum from the coherent (reproducible) portion of the pulse would be strongly peaked at 80N

Hz, and would be $8 \text{ Hz} \times 8 \text{ sec} = 640$ times larger at the peak than the spectrum from a single pulse. The white background has 1% of the total energy of the pulses, or 6.4 times the energy of a single pulse, so the field is on average $(6.4)^{1/2} \approx 2.5$ times larger than the single pulse spectrum at an arbitrary frequency. Thus the peak amplitude is $640/2.5 \approx 48$ dB above the baseline, and the ratio (power at peak frequencies)/(power far from peak frequencies) scales as the first power of the number of pulses. This baseline value, corresponding to return from the pulse jitter, exceeds the 60 dB selectivity discussed in the last paragraph, and would in fact provide the limitation to sensitivity. With an abruptly terminated periodic waveform the two contributions would be comparable after 40 pulses, or 0.5 sec; jitter effects will dominate even earlier with other models for the truncation.

4. The **optimum pulse spectrum** can be predicted from a number of simple considerations. First of all, it is important to realize that radar cross-section is a strong function of wavelength. The key parameter is the ratio of a characteristic dimension of the scatterer (L) to the wavelength of the incident field (λ). When the characteristic dimensions of the scatterer are comparable to or larger than the wavelength calculating cross sections is extremely difficult. However, since $\lambda = 10$ meters at the upper end of the HF range (3-30MHz), and 100 meters at the lower end, we generally have $L/\lambda < 1$. When $L/\lambda \ll 1$ scattering behavior is Rayleigh-like; this is of course the same basic phenomenon which makes the sky blue. In this limit the radar cross section scales as the fourth power of the scatterer dimension (see Figure 7).

Thus most of the energy in the "start pulse" described in equation [1] is wasted, because the low frequency components generate weak reflections. It would be sensible to alter the total field profile to enhance the higher frequency components to maximize return. In fact, this turns out to be a natural consequence of reflections and the "stop pulses", because they have the opposite sign. Consider, as a simple approximation, an electric field profile consisting of a "start pulse" and an equal amplitude inverted reflection from ground 20 ft away. The round trip time to the ground is 40 ns, so the net effect at vertical elevation would be a pulse which looks roughly like a single cycle of a square wave. This creates an effective carrier frequency of 12.5 MHz, and concentrates energy in the HF frequency region at the expense of dc components. At other elevations there will be partial electric field cancellation, as the effective reflection delay $\Delta t = (2h/c) \sin \chi$ is shorter. The field cancellation reduces the total useful radiated energy, but the shorter time delay gives an even stronger concentration in the higher frequency regions which are of actual interest.

Based on these arguments about waveforms and CPB constraints, we can predict that the optimum performance might be achieved by a waveform which permitted a few tens of nearly reproducible pulses to hit the target in a time less than 1 sec, and then was scanned to other

positions. A longer dwell time causes the normal square-root-of-time improvement in detection of coherent versus incoherent signals.

IV. Alternatives for HF Radar Detection

The results of actual experimental measurements of CPB profiles are classified. However, it is safe to conclude that electric fields this large (and their reflections) are readily detected with relatively small antennas. For most applications, then, the very large receive antenna in Figure 2 can be replaced with much more portable devices.

It becomes sensible to consider novel detectors and preamplifiers in the limit of extremely small signals. One interesting alternative is the use of r.f. amplifiers based on superconducting quantum interference devices (SQUIDS). These devices normally have a very restricted frequency range (up to a few kHz) but recent work has produced extremely low noise amplifiers at frequencies up to 300 MHz with d.c. SQUIDS⁵. In addition, SQUID operation has recently been verified at 77K in $\text{YBa}_2\text{Cu}_3\text{O}_y$ ⁶, and applications of these high- T_c superconductors to r.f. preamplifiers seem likely.

V. Radar Countermeasures

There are four, fundamentally different approaches to radar cross section reduction: shape alterations, addition of absorbing materials, passive spoofing, active spoofing. The first two ideas can be successfully employed in the UHF and microwave regions, but are not successful at HF. As noted earlier when $L/\lambda \ll 1$ scattering behavior is Rayleigh-like, and in this limit the radar cross section is nearly shape independent, but scales as the fourth power of a characteristic scatterer dimension (see Figure 7). Smoothing the surface does very little; coating with dielectrics is ineffective until the thickness of the coating approaches $\lambda/4$ (the principle behind such coating schemes is identical to the design of antireflection coatings for optics); and radar absorbing coatings tend to be effective only over small fractional bandwidths. Passive spoofing schemes, such as designing a resonant cavity at the right frequency to cancel reflections in one direction, are narrowband and require too much size.

Active spoofing schemes would require detection of this high power pulse, generation of a replica within a few nanoseconds, and amplification to extraordinarily high peak powers. This is

likely to be extremely difficult, and in the best of circumstances will only achieve significant cancellation in a single direction. Thus if the receiver position is not accurately known this countermeasure is nearly useless.

VI. Conclusions

We have tried to present a qualitative overview of the advantages and limitations of wideband radar generation and detection, with particular emphasis on the uses of charged particle beams as HF radar transmitters. We conclude on very general grounds that the spatial and temporal profile of typical fields induced by CPB propagation are well matched to the requirements for these applications, and thus the advantages of concealability and resistance to radar countermeasures make further exploration of this approach worthwhile.

I wish to thank Lee Buchanan of DARPA's Directed Energy Office for pointing out this application and providing helpful background material, Ed Lyon of SRI for a briefing on OTH radar systems, and many of the Defense Science Study Group members for useful conversations as this topic developed.

Figure Captions

Figure 1. Transmit Antenna for the East Coast Over-the Horizon Backscatter (OTH-B) radar system. Figure courtesy of Dr. Edwin Lyon, SRI.

Figure 2. Receive Antenna for the East Coast Over-the Horizon Backscatter (OTH-B) radar system. Figure courtesy of Dr. Edwin Lyon, SRI.

Figure 3. Temporal profile of a typical charged particle beam pulse. Adapted from ref. [1]

Figure 4. Geometry of a charged particle beam and the accelerator exit port. Adapted from reference [1]

Figure 5. Normalized electric field patterns for the "start pulse" (equation (1)) assuming the geometry in Figure 4.

Figure 6. Electric field contours for the specific case $\beta=0.95$, in units of peak field (volts per meter) over peak current (amps). Thus a 10 kA peak current beam gives a 10kV/m field on the contour labeled "1". Adapted from ref. [1].

Figure 7. Generalized Radar Cross Section curve, showing the characteristic λ^4 cross section dependence in the Rayleigh limit $L/\lambda \ll 1$.

¹The exact numbers are classified. This unclassified estimate comes from "Electromagnetic Effects of Charged-Particle Beams", by Kendall F. Casey of JAYCOR, 39650 Liberty Street, Suite 320, Fremont, CA 94538

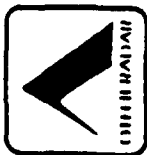
²The convention used here, $1 \text{ Avis (Av)} = 1 \text{ rad/sec}$, was originated by J. S. Waugh at MIT.

³N. J. Carron, "Exterior Beam Self Fields", Mission Research Corporation, P. O. Drawer 719, Santa Barbara, CA 93102

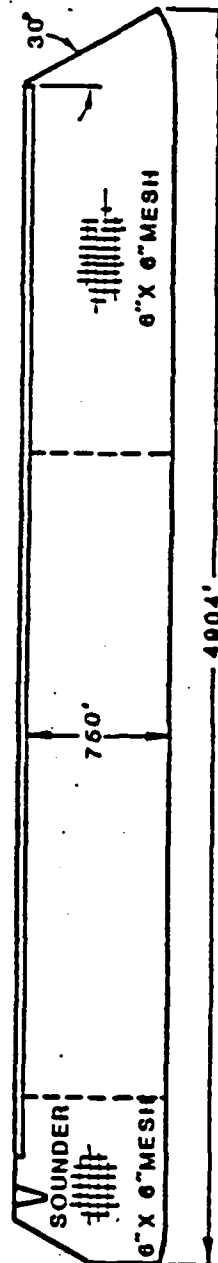
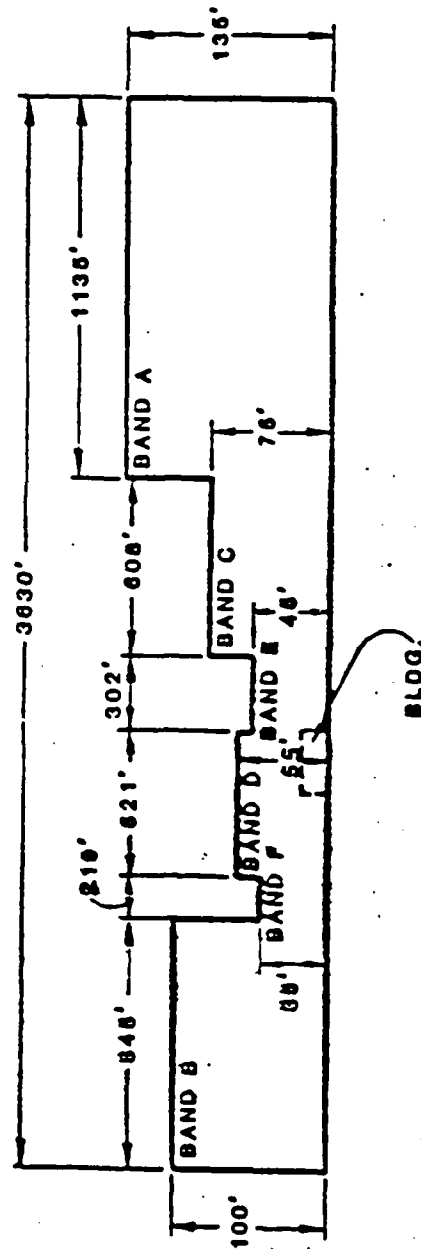
⁴I. LaHaie, R. O. Harger, S. R. Robinson and J. M. Miller, "An Evaluation of Nonsinusoidal Radar Techniques", Environmental Research Institute of Michigan, June 1985

⁵C. Hilbert, J. Clarke, T. Sleator, E. L. Hahn, Appl. Phys. Lett. 47, 637 (1985); T. Sleator, E. L. Hahn, M. B. Heaney, C. Hilbert, and J. Clarke, Phys. Rev. Lett. 57, 2756 (1986)

⁶C. M. Pegrum, G. B. Donaldson, A. H. Carr and A. Hendry, Appl. Phys. Lett. 51, 1364(1987)



TRANSMIT ANTENNA-PLAN VIEW

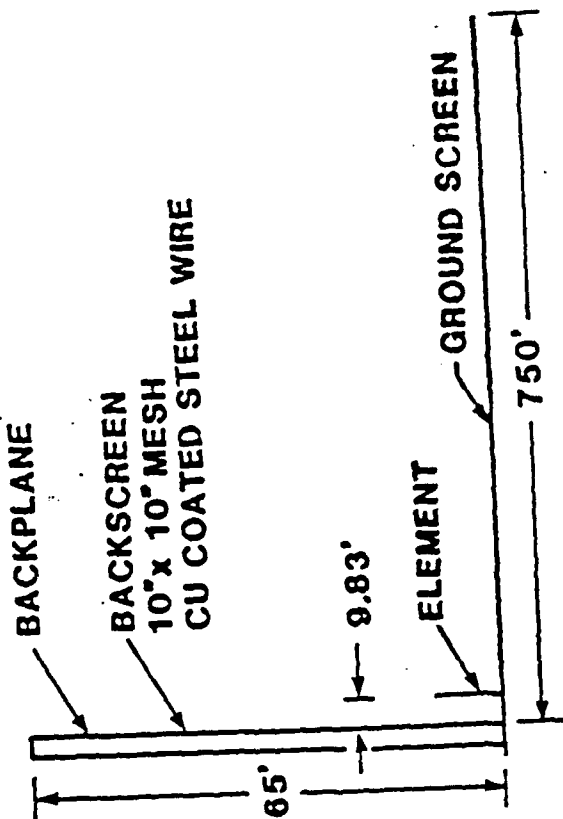
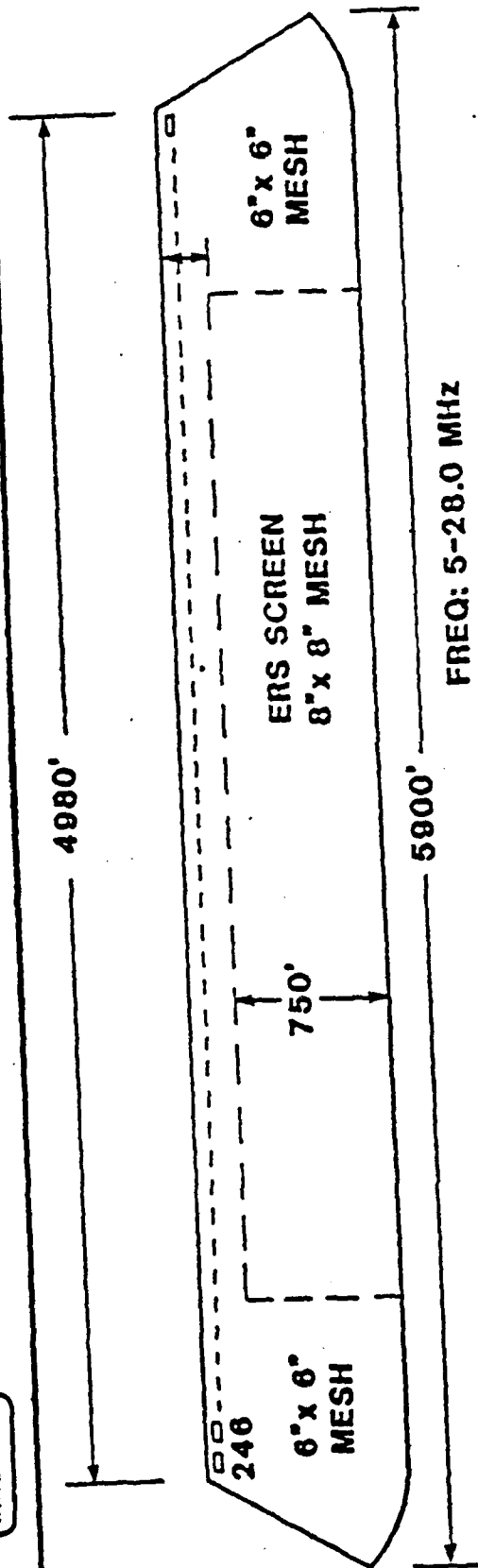


BAND ELEMENT SPACING (FT)

BAND	ELEMENT SPACING (FT)
A	90.49
B	67.13
C	49.80
D	36.04
E	27.41
F	20.33



RECEIVE ANTENNA PLAN VIEW



NEL/BAND: 82

BAND-3

HIGH 15.0-28.0 MHz

MID 9.79-15.0 MHz

LOW 5-9.79 MHz

ELEMENT SEQUENCE

HIGH 83-164 ADJACENT

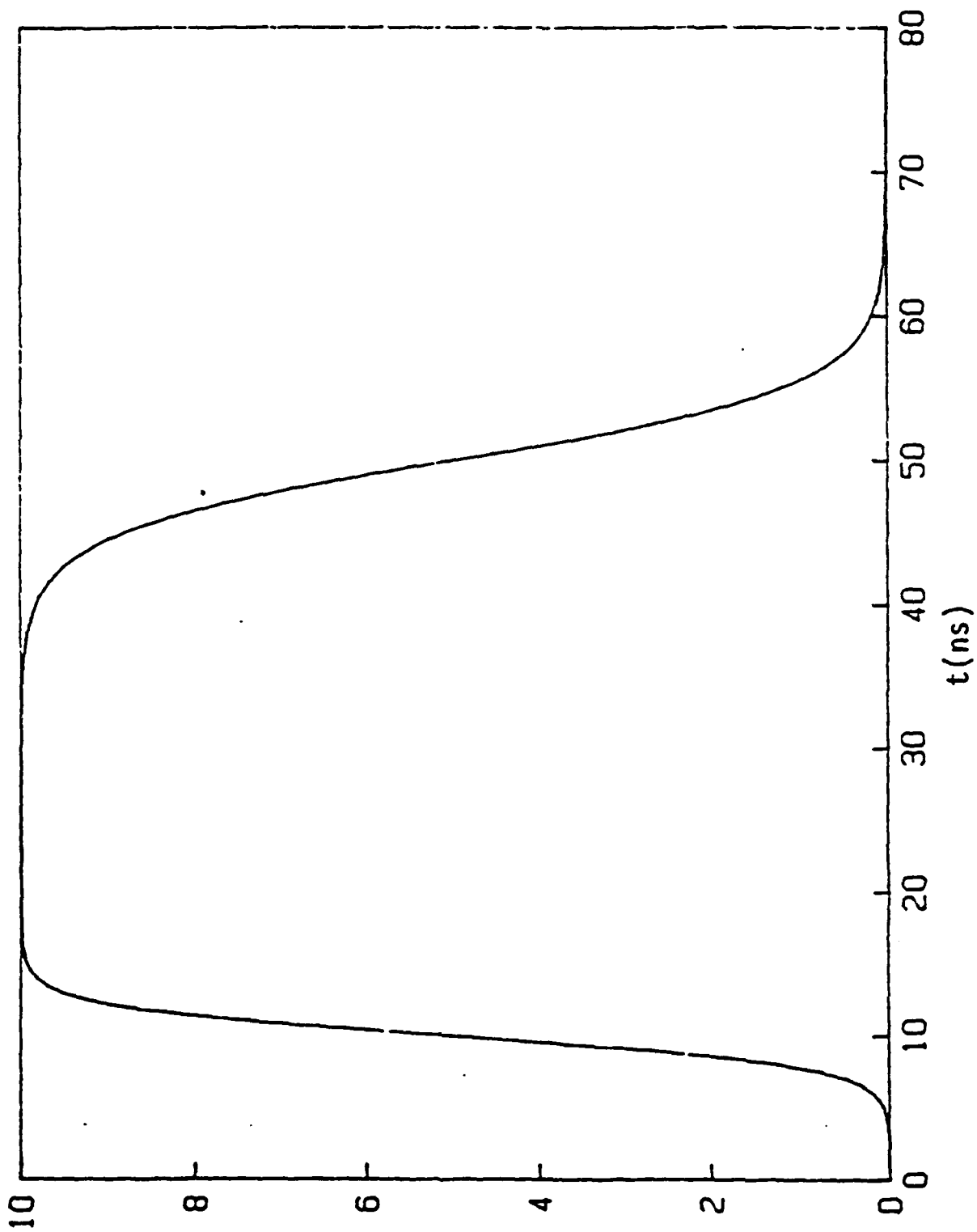
MID 42-204 EVEN NO's

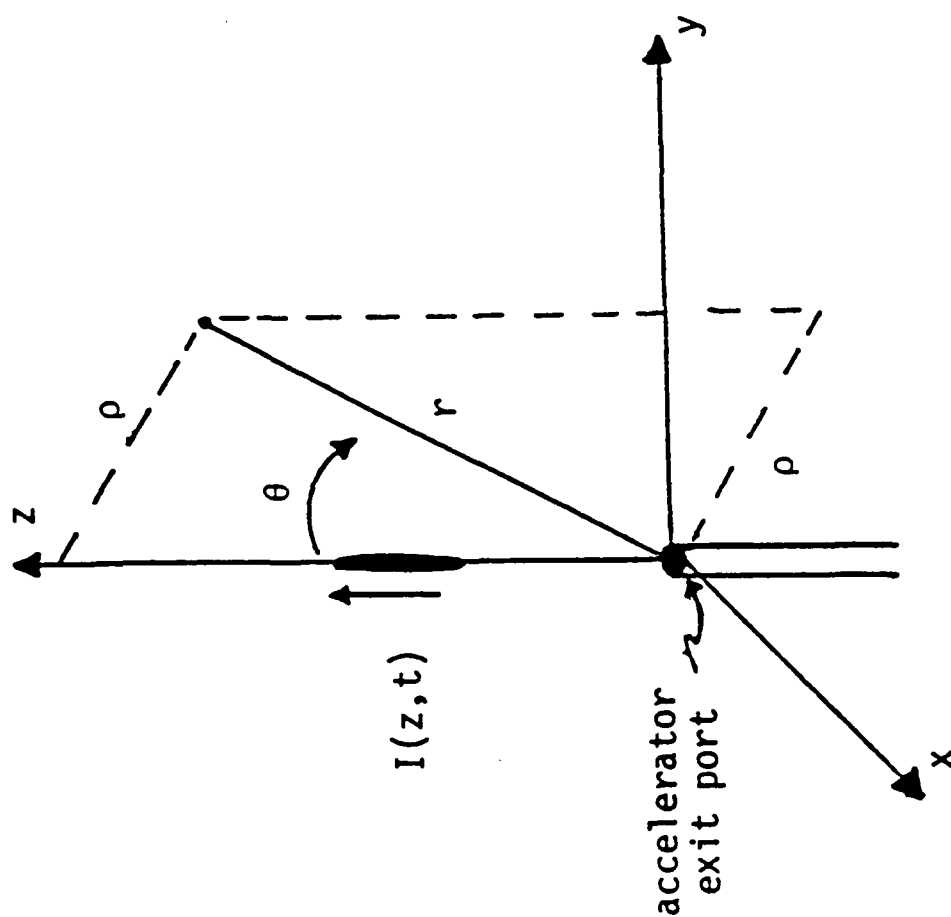
LOW 2-246 EVERY 3RD ELE

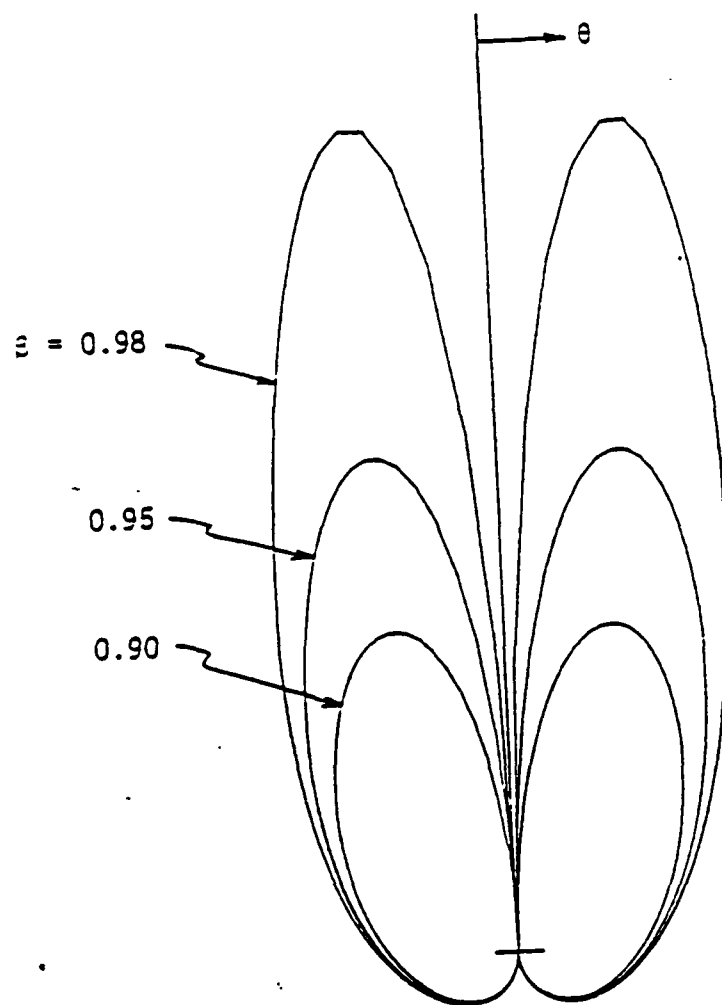
DUMMY ELEMENTS 80

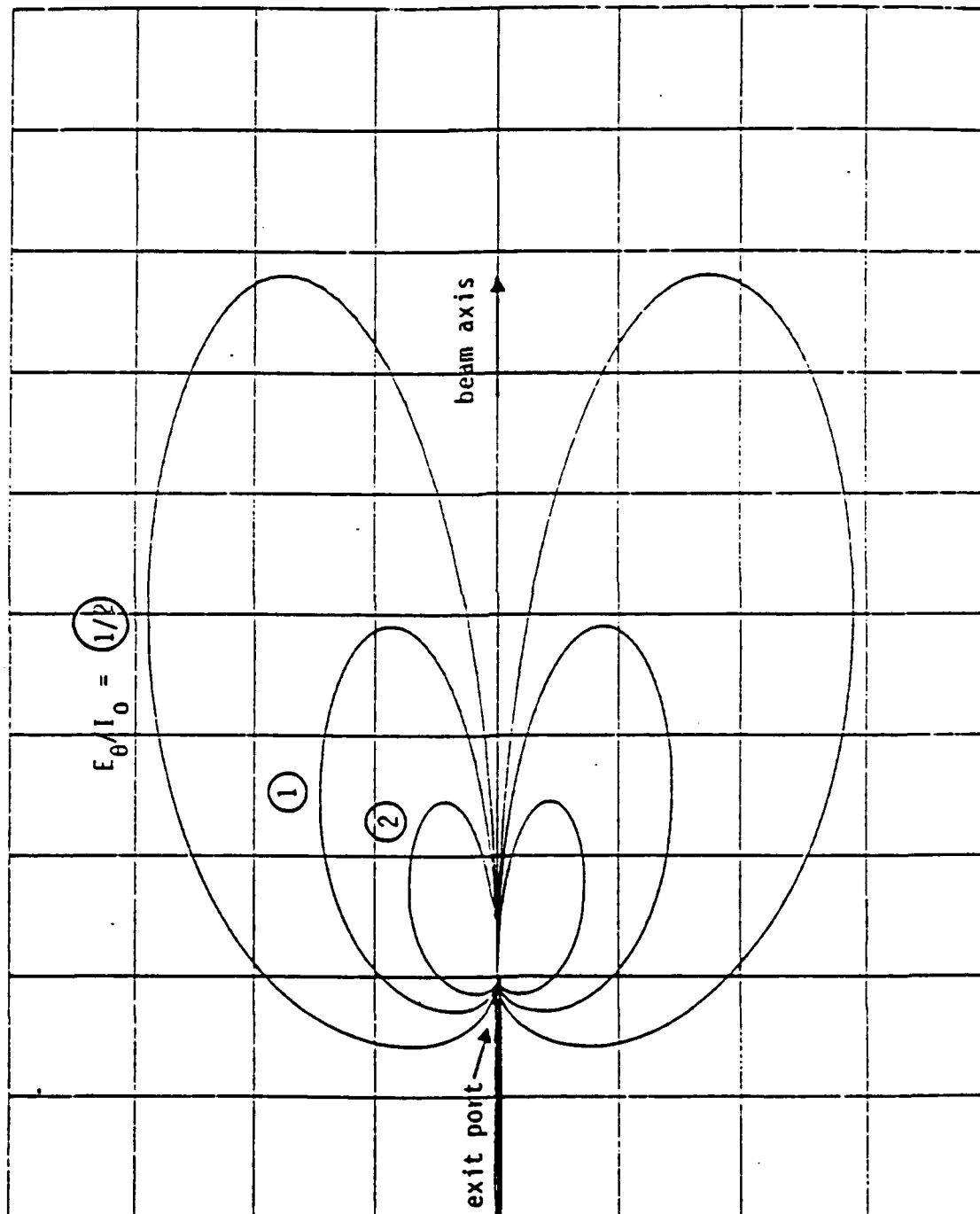
$I(t)$
(kA)

III-98

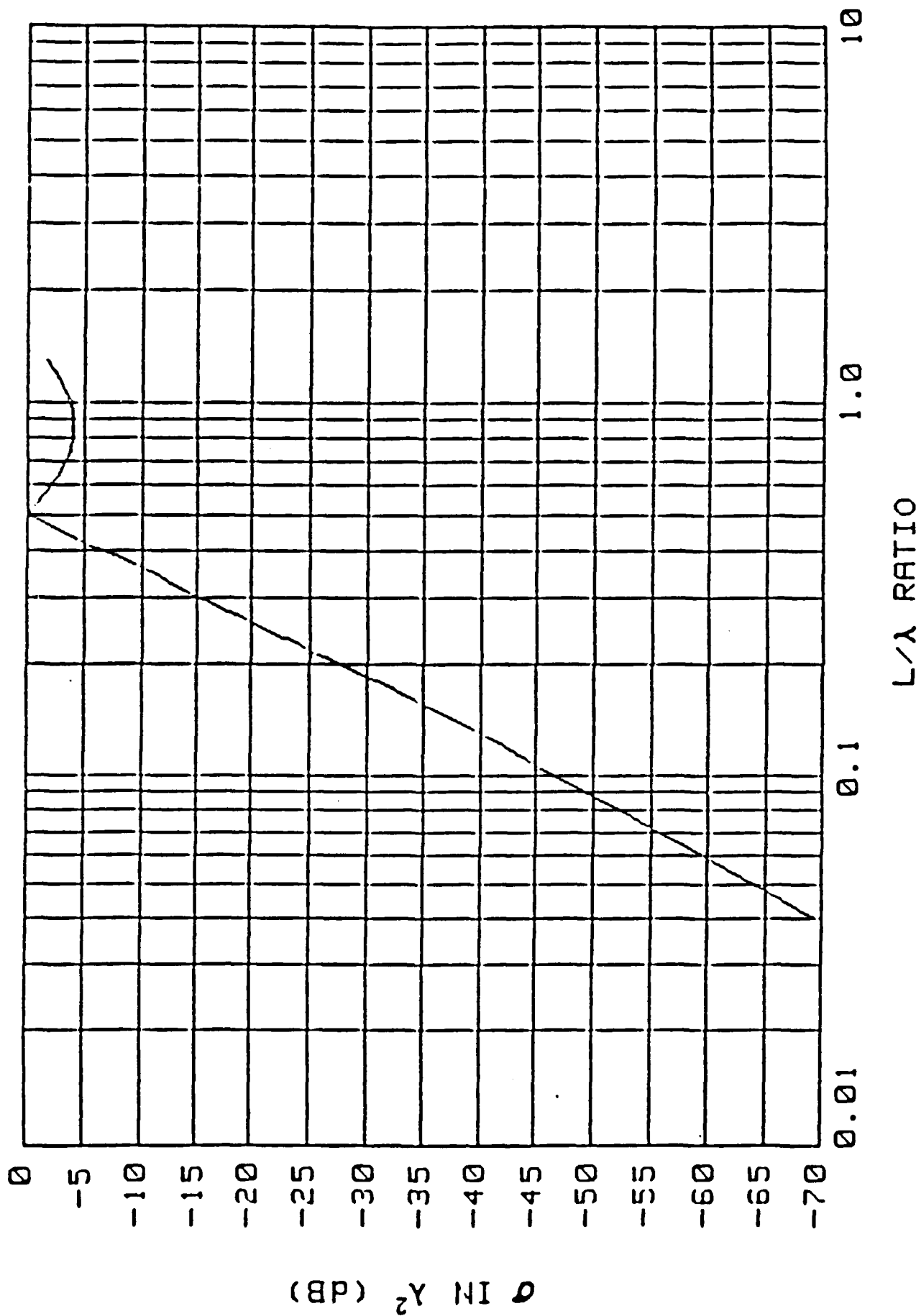








GENERALIZED RCS CURVE



**D. POPEYE - AN EXPENDABLE BATTLEFIELD
SURVEILLANCE SYSTEM**

AUTHORS:

Russ Caflisch
Steve Case
Deborah Joseph
Bruce Hajek
Randy Katz

Steve Koonin
Tom Rosenbaum
Steve Sibener
Warren Warren
Hugh Woodin

CONTRIBUTORS:

Col. William Wilson
Commander, 177th Armored Brigade, Ft. Irwin

LTC Peter Manza
Commander, 63rd Armor OPFOR, Ft. Irwin

Jim O'Brian
Fort Huachuca

Richard Schwartz
David Hardison
IDA

Dr. William Isler
DARPA

OUTLINE

- I. Introduction - The Problem and Goals
- II. Popeye - A Proposed Solution
- III. Technology
 - A. System Performance Specifications
 - B. Dynamics of Artillery Launch
 - 1. Kinematics
 - 2. Midflight Corrections to alter Kinematics
 - 3. Launch Impulse
 - C. System Design
 - 1. Overview
 - 2. Thermal Imaging with a single HgCdTe Detector
 - a. Blackbody Radiation
 - b. Lens System and Detector Design
 - D. Projectile Angular Velocity Determination
 - E. Projectile Rate of Descent and Frame Time
 - F. Data Transmission
 - G. Synchronization and Framing
 - H. Summary
- IV. Uses for Data
- V. Cost Issues
- VI. Conclusions

I. Introduction - The Problem and the Goals

The Defense Science Study Group toured the Army National Training Center on June 13-14, 1988, and witnessed a simulated dawn tank battle between American and Opposing Forces (OPFOR) battalions. It quickly became apparent that a major difficulty in many battle scenarios was the inability to get up-to-date position information. This data is invaluable in planning for an attack, in eliminating entrenched tanks, and for making mid-attack corrections. Pre-battle artillery attacks relied heavily on advanced intelligence teams, as shown in Figure 1, whose members are vulnerable and often provide dated positions. At present, the cost of a sophisticated observation platform such as a helicopter, airplane, or drone is sufficiently high that ground troops are the primary data gathering means.

Without accurate position data, blue-on-blue fratricide attacks were not uncommon; low visibility made positive identification difficult. Individual tank commanders did not know their own locations, or the locations of opposing tanks or dismounted attacking troops. An inexpensive, near real-time method to provide position data under various conditions-day/night, moderate to heavy smoke-would appear to be extremely useful.

II. Popeye - A Proposed Solution

We propose here an artillery launched sensor capable of large area remote surveillance that is buildable with current technology and is affordable.

The system proposed consists basically of a smart artillery shell. The shell contains relatively simple optics and in its simplest implementation, a *single point detector* for each portion of the spectrum in which images are desired. In this report, we restrict our calculations to the case of a single detector geometry. We analyze the light budget for the sensor and the required projectile kinematics. We show that current commercially available optics and electronic components can generate a thermal image of a 2.5 km diameter area with resolution of 1 meter at the edges (and proportionately better at the center) in a few seconds.

The concepts we will discuss here evolved during the months after our visit to the NTC but use ideas and systems concepts we have gathered during our various visits and study sessions. Because of the very limited time available to this study, we must inject the disclaimer that most of the calculations that follow are performed to test the feasibility of the project on scientific grounds.

In the discussions that lead up to this project we sought a very simple sensor concept. A schematic of our design is shown in Figure 2. Here a rotating projectile drops into the surveillance area. The projectile contains a lens and a *single* detector element such that the optical axis is oriented 30° off axis from the direction of forward travel of the shell. At any instant, the sensor observes a zone of size Δr on the ground. As the shell spins, the sensor traces out a circle on the ground. As the shell falls, the diameter of the circle decreases so that an entire two dimensional picture is built up as shown in Figure 3 and transmitted to a remote location.

Several possible trajectories we considered are shown in Figure 4. The last system involving a slow descent and slow spin rate was most compatible with our resolution and data rate goals and will be analyzed here.

III. Technology

The following section shows our calculations for system feasibility. Full sensor design equations are given so that as assumptions in our model are changed, performance characteristics can be recalculated. We show the design only for an IR detector so that the system can afford nighttime and thru-smoke visibility. The addition of a silicon detector for visible light performance is a straightforward extension. We address several issues involved in the full sensor system design including the kinematics of projectile motion in addition to optical system design.

III.A. System Performance Specifications

The performance specifications for our IR detection system are shown in Table I.

Table I Sensor System Specifications

Ground Coverage:	2.5 KM Diameter Circle
Ground Resolution:	1 M
Frame Time:	Less than 30 seconds
Temperature Resolution:	1°K (Water cooled tank engine is 50° -30° = 20°C = 4 bits above ambient)
Projectile Diameter:	155 mm
Lens Diameter:	100 mm

III.B. Dynamics of Artillery Launch

III.B.1. Kinematics

The dynamics for an artillery launched 155mm projectile are in Table II. Because the forward travel per revolution is 3 meters we decided not to pursue the overflight trajectory in Figure 4. This is because the resolution along the direction of travel would be 3 meters unless a multi-detection geometry were employed.

Table II. Kinematics: 155 Projectile

Muzzle Velocity:	900 M / sec
Rotation Rate:	300 Rev / sec
Forward Travel/Revolution:	3 M
Range:	20 KM
Endpoint Accuracy:	200 M

III.B.2. Midflight corrections to alter Kinematics

Because the descent and rotation rates for proper projectile flight may not be ideally matched to desired rotation rates, we briefly investigated if technologies existed to alter projectile kinematics after it flew to the desired surveillance area. We found that two systems: a Rotation Brake (RAID) and a Vortex Ring Parachute can be used to control descent and rotation rates quite accurately. We thus could achieve a desired descent as in the last drawing in Figure 4.

III.B.3. Launch Impulse

Using the simple formula

$$v^2 = 2 a s$$

with

v = muzzle velocity = 900 m/sec

s = cannon barrel length = 3 m

g = acceleration of gravity = 9.8 m/sec²

the acceleration of the projectile during launch is

$$a = v^2 / 2s = 135,000 \text{ m/sec}^2 \\ \approx 10^4 \text{ g's}$$

We were concerned with whether sensors and communications equipment could survive this acceleration. We found that an artillery launched CCD camera had been developed and that a laser guided artillery shell has been deployed indicating that electronic and optical systems could be cannon launched.

III.C. System Design

III.C.1. Overview

The proposed method of descent is shown at the bottom of Figure 4. This coupled with the system specifications in Table I allow us to establish the design parameters in the following section. The calculation proceeds in several parts. First we must calculate the flux of useful thermal radiation at the detector from each 1 m² spot on the surface. We then must calculate the effects of finite lens diameter to determine the minimum size detector possible, since this minimizes detector noise. We then use the detectivity of HgCdTe to calculate the maximum bandwidth consistent with a 1:1 signal-to-noise ratio for the least significant bit in our 4-bit intensity image. Finally, we use this bandwidth to determine the constraints on projectile angular velocity, rate of descent, and total image time. We conclude that these design parameters can be met with a single point sensor.

We initially assume that the projectile plus drag parachute will fall at constant velocity, will not drift, and will have constant angular velocity.

III.C.2. Thermal Imaging with a single HgCdTe point detector

III.C.2.a. Blackbody Radiation

The Planck blackbody radiation law states that M, the spectral radiant exitance from a blackbody at temperature T, is $2\pi hc^2 / \lambda^5 [e^{hc/\lambda kT} - 1]$. The units we will use are watts per m² of surface area, per micron of optical wavelength. The change in emitted power per unit temperature difference is given by:

$$\frac{dM}{dT} = \frac{2\pi hc^2}{\lambda^5 [e^{hc/\lambda kT} - 1]^2} e^{hc/\lambda kT} \frac{hc}{\lambda kT} \frac{1}{T}$$

For convenience let $u = hc/\lambda kT$:

$$\frac{dM}{dT} = \frac{2\pi hc^2}{\lambda^5 [e^u - 1]^2} e^u u \frac{1}{T}$$

Now if $\lambda = 10 \mu\text{m}$ and $t = 300 \text{ K}$,

$$u = \frac{(6.6 \times 10^{-34} \text{ W sec}^2)(3 \times 10^8) \text{ m/sec}}{(10^{-5}) \text{ m} (1.38 \times 10^{-23}) \text{ W sec/}^\circ\text{K} (300^\circ\text{K})} = 4.78$$

$$\frac{dM}{dT} = \frac{2\pi (6.6 \times 10^{-34})(3 \times 10^8)^2}{(10^{-5})^5 [e^{4.78} - 1]^2} e^{4.78} \frac{4.78}{300}$$

$$= 4.97 \times 10^5 \text{ W/(m}^2 \text{ of surface)(m of wavelength)(}^\circ\text{K)}$$

$$= 0.50 \text{ W/(m}^2 \text{ of surface)(}\mu\text{m of wavelength)(}^\circ\text{K)}$$

Assume that the desired spot size $A_{\text{pixel}} = (\Delta x)(\Delta r) = 1 \text{ m}^2$ (see Figure 3), the detection spectral width $\Delta\lambda = 2 \mu\text{m}$ ($\lambda = 9 \rightarrow 11 \mu\text{m}$), and the desired temperature resolution dT is 1°K . Then

$$dM = (0.5) \cdot A_{\text{pixel}} \cdot \Delta\lambda \cdot \Delta T = (0.5) \cdot 1 \cdot 2 \cdot 1 = 1 \text{ watt}$$

This is the difference in emitted power (in the 9-11 μm band) from a 1 m^2 blackbody whose temperature is 1°K different from its neighboring 1 m^2 pixel.

Since the earth (or a tank engine) is not a perfect blackbody, we must multiply by the emissivity of the engine surface. A conservative estimate would be $\epsilon = 0.2$. This implies that $0.2 \times 1 = 0.2$ watts emitted by the surface must be detectable by our sensor system. This power is emitted into 2π steradians; the fraction of the light incident on the detector is found by calculating the usable solid angle Ω , which in this case is the lens area divided by the square of the distance, or $(\pi \times (5 \text{ cm})^2) / (2.5 \times 10^5 \text{ cm})^2 = 1.26 \times 10^{-9}$ steradians. The flux ϕ on the detector is then given by

$$\phi = (0.2 \text{ W}) \cdot \Omega \cdot \cos 30^\circ / 2\pi = 35 \text{ picowatts}$$

The cosine factor comes from the geometry in Figure 2; the source is tilted by 30° with respect to the detector. This minimum flux should be equal to the noise equivalent power (NEP) for our detector to give 1:1 signal-to-noise for the least significant bit.

III.C.2.b. Optical System and Detector Design

Assume that the detector is angled 30° from the vertical, as in Figure 2. We then have

$$r_o = \frac{2.5 \text{ km}}{2} = 1.25 \text{ km}$$

$$z_o = \frac{r_o}{\tan \theta} = \frac{1.25 \text{ km}}{\tan 30^\circ} = 2.17 \text{ km}$$

If the lens has a 10 cm diameter and 20 cm focal length (f/2 lens, which is fast but buildable), the angular resolution is

$$\Delta\theta = \frac{2.88\lambda}{D} = \frac{2.88 (10\mu)}{10 \text{ cm } 10^4 \frac{\mu}{\text{cm}}} = 2.88 \times 10^{-4}$$

$$\text{The object distance at the beginning of the scan} = \frac{Z_o}{\cos \theta} = \frac{2.17}{\cos 30^\circ} = 2.5 \text{ km}$$

$$\begin{aligned} \text{The diffraction limited ground resolution is then } (\Delta\theta) (Z_{\text{obj}}) &= (2.9 \times 10^{-4})(2.5 \times 10^3) \\ &= 0.7 \text{ m} \end{aligned}$$

Therefore this lens can just resolve a 1 m patch on the earth, and is satisfactory. The detector will be located 1 focal length after lens (since the object is effectively at ∞). For a 20 cm focal length, the diffraction limited spot size is

$$\Delta x_d = 2\lambda f/D = 2(10\mu)(20\text{cm})/(10\text{cm}) = 40 \mu\text{m}$$

The detector can be 50% larger than this and still obtain the 1 m ground resolution. This implies a detector size of $60 \mu\text{m}$, which is commercially available. From spec sheets, the detectivity D^* for a HgCdTe detector @ 10μ , cooled to 77K, is $10^{11} \text{ cm (Hz)}^{1/2}/\text{W}$. The noise equivalent power $\text{NEP} = \emptyset = 35\text{pW}$ is related to D^* via the detector area A_d and system bandwidth Δf

$$D^* = \frac{\sqrt{A_d \Delta f}}{\text{NEP}}$$

$$\begin{aligned} \Delta f &= \frac{(D^* \text{NEP})^2}{A_d} = \frac{[(10^{11})(35 \times 10^{-12})]^2 \text{ cm}^2 \text{ Hz}}{\left[(60 \mu\text{m}) \left(10^{-4} \frac{\text{cm}}{\mu\text{m}} \right) \right]^2} \\ &= 333 \text{ kHz} \end{aligned}$$

In other words, the sensor must dwell for a time on the order of $1/\Delta f = 3\mu s$ on the $1m^2$ pixel in order to obtain adequate exposure.

III.D. Projectile Angular Velocity

$$\begin{aligned}\omega_o &= \frac{\Delta\theta}{\Delta t} = \Delta\theta \Delta f = \frac{\Delta x}{r_o} \Delta f = \frac{1m}{(1.25 \times 10^3 m)} 333 \times 10^3 \frac{1}{sec} \\ &= 266 \text{ radians per second} \\ v_o &= \omega_o/2\pi = 42 \text{ revolutions per second}\end{aligned}$$

III.E. Projectile Rate of Decent and Frame Time

After 1 revolution, which takes time

$$T = \frac{1}{v_o} = \frac{2\pi r_o}{\Delta x \Delta f},$$

the projectile should have dropped in height so that the observation spot has moved over by $\Delta r = 1 \text{ meter} = \Delta x$, where $\Delta r = \Delta z \tan \theta$. The vertical velocity is therefore

$$V_T = \Delta z/T = \left(\Delta r/r_o\right) (\Delta x \Delta f / 2\pi \tan \theta) = 98 \text{ m/sec}$$

Our frame time is then

$$T_{total} = z_o/V_T = (2.17 \times 10^3 m) / (98 \text{ m/sec}) = 23 \text{ sec}.$$

III.F. Data Transmission

The relatively low frequency 300 kHz data rate allows over-the-horizon communications using a trailing wire antenna behind the projectile. We were told by a DSSG mentor that one of the most difficult problems in this entire project may be establishing a secure data transmission means. Toward resolving this problem, we also investigated the use of fiber optic communication and found that mortar launch missiles (FOGM program) are technically capable of communicating through a trailing fiber optic cable.

III.G. Synchronization And Framing

Several methods were discussed to synchronize the scans from the rotating projectile. We concerned ourselves here with the possibility of non-uniform angular velocity during descent.

The ideas discussed included using a gyroscope; using a directional antenna to send a sync signal; or correlating scene data in post observation data reduction.

The last technique, synchronization by scene correlation, is illustrated in Figure 5. In Figure 5a, the spiral scan over a road is shown. In Figure 5b, a plot of the sensor signal vs. time shows the signature of the road on two successive rotations of the sensor. In Figure 5c, we perform a correlation between tracks in the scan pattern to properly register the track-to-track data. This system could reduce the on-projectile hardware but is risky if large areas of featureless terrain are encountered.

III.H. Summary

The summary of our analysis is listed in Table III. It appears that the sensing system could be built with currently available components. Designing a system to include visible wavelength optical images could also be done by adding a single element silicon detector and most likely an additional lens. In general, visible light detection is easier than IR so that we have worked out the more difficult case. Incorporating both visible and IR detectors into the same projectile would add capabilities with modest increased technical difficulty.

Table III Proposed System Summary

Detector:

Assume HgCdTe

**Cooled to 77° K (airborne dewar may not be required since
airborne lifetime is short)**

60 x 60 μ m detector area

NEP--Noise equals signal from 1°K temperature different object

**Bandwidth: using sensitivity (Detectivity) of commercially
available detector, data rate is 333 kHz**

Kinematics for Sensing:

Rotation Rate:	42 Rev./ sec
Descent Rate:	98 m / sec
Descent Time:	23 sec
Scan Area:	2.5 km diam. circle

IV. Uses For Data

Multiple uses for the data from the proposed sensor have been identified. IR visualization of the scene in a pre-dawn timeframe could provide timely data on red team tank locations. Through sequential observations, movements of both red and blue team forces would be monitored, yielding better blue force coordination. Finally, the broad field, moderate resolution sensor proposed here could be useful to direct SADARM munitions (which have shorter range sensing capabilities) against entrenched tanks.

V. Cost Issues

We concerned ourselves briefly with the cost of some of the components in the proposed system. Commercial prices were found for detectors and lenses. IDA personnel gave us estimated prices for some system components and for munitions. Some of these are listed in Table IV. The systems prices were used as a reference. The component prices were found to ensure that some of the critical components in the system would not be so expensive as to make the system impractical from a cost perspective.

Table IV Costs

Components:

Radar Altimeter	\$300
Gyroscope	1,000
HgCdTe Detector	1,200
Ge Lens (5 cm diameter)	400

Cost References:

155 mm shell with plain explosives	\$500
155 mm shell with 80 bomblets	1000
155 mm shell with 2 SADARM submunitions	6000
Copperhead Guided Missile	35,000
Drone	600,000

VI. Conclusions

Within the limited time available to pursue this research, the Popeye concept appears unique and we have not uncovered any scientific objections to its feasibility. Of course, none of the DSSG members are weapons designers so we cannot address the practicality of the implementation of this approach.

From our site visit to the NTC, it appears that there is a need for this type of large area surveillance system and we hope that this brief report may serve as an outline for further research in this area.

The numerical values of parameters chosen to use in our design equations are estimates. Their intention was to calculate the feasibility of this approach. A reviewer of this manuscript has suggested that making the spatial resolution equal to 0.5 meter so that a 4 line pair image of a tank is available, could allow red vs. blue tank discrimination. This would be an interesting new development and is gratifying since our initial specifications were very close to this value. Again we must reiterate that the period available for this research was fairly short and we hope that it serves to stimulate future ideas for our sponsors.

E. SOME PROMISING DIRECTIONS IN APPLIED MATHEMATICS

*Russel Caflisch
Bruce Hajek
Deborah Joseph
Randy Katz
Steven Koonin
Philip Marcus
Daniel Stein
Warren Warren
Hugh Woodin*

Defense Science Study Group
Institute for Defense Analyses

1. Introduction

Many new opportunities and applications for applied mathematics have been created by the computerization of science and technology. Increased speed and memory and decreased cost of computers are making them convenient and effective tools for mathematical experimentation and scientific prediction. The goals of this report from the Defense Science Study Group are to identify several promising current developments in applied mathematics, to point out some applications that are in need of additional mathematical tools and to suggest some new technical directions. The report focuses on two subjects: filtering and transform methods, and numerical methods that employ randomness. A summary is given of a recent report on future directions in control theory.

2. Filtering and Transform Methods

Filtering and transform methods are used in a variety of scientific applications, including signal processing, imaging, data compression and scientific computation. Mathematical developments, such as the discovery of the Fast Fourier Transform (FFT), have had a tremendous effect on the efficiency and range of applications of these methods, and future developments should have an equally great impact. Two topics are discussed here: the new "wavelet" transform method and the need for more effective methods of parametric filtering.

Fourier Transform and Wavelets

The Fourier transform is the most important and useful transform method and has some clear advantages over any alternatives. First it is a simple method to use and understand. Second it is a natural method, since the principal modes of many problems are naturally expressed in the harmonic basis. For example the cochlear membrane of the inner ear has been found to decompose sounds into their Fourier components. Finally there is a fast numerical method, the FFT, for computing the Fourier transform and its inverse.

However the Fourier method also has some strong disadvantages. For many problems Fourier components do not well represent the principal modes. Each Fourier mode gives a point location in frequency space; whereas often the widths of peaks and temporal decay properties are equally important. Second the Fourier modes are not localized in space. They represent an extreme case for the uncertainty principle, since their frequency is determined exactly and their position is completely undetermined, whereas for many problems it is desirable to have some compromise between frequency and spatial information. For example in image processing or scientific computation, spatial control over the modes of the filter would allow refined representation in regions of complex structure and coarse representation elsewhere, with an overall saving in the number of required modes.

An alternative transform method that combines many of the advantages of the Fourier transform without the corresponding disadvantages is that of "wavelets". The wavelet basis consists of a doubly indexed set of functions

$$\phi_{nk}(x) = \phi(2^n x - k).$$

formed from a single, specially chosen function ϕ . The shift k picks the spatial location of the function; the scale factor 2^n picks the refinement level, which is related to the wavenumber. The function ϕ is chosen to be localized in both position and wavenumber. In addition, special choices of ϕ have been found [4,5,15] for which distinct modes ϕ_{nk} contain distinct information; i.e. ϕ_{nk} are orthogonal, satisfying

$$\int \phi_{nk} \phi_{ml} dx = 0 \quad \text{if } (n,k) \neq (m,l)$$

This means that maximal data compression is achieved with wavelets as a basis and that the expression of a function in terms of the wavelet basis is simple. For any function f

$$f(x) = \sum_{n,k=0}^{\infty} \alpha_{nk} \phi_{nk}(x)$$

in which

$$\alpha_{nk} = \int f(x) \phi_{nk}(x) dx$$

In these integrals the functions f and ϕ are considered to be periodic and the integrals are over one period. There are analogous formulas for the whole line case.

Figure 1 shows one such choice of a wavelet function ϕ and its Fourier transform. An example of its use is shown in figure 2 for image processing in two dimensions. The figure at the left is decomposed into its wavelet components at the right. At each level there are three images, one corresponding to a fine vertical scale and a coarse horizontal scale, the second to a fine horizontal scale and a coarse vertical scale and the third to fine horizontal and vertical scales. Each of these images has a quarter as many pixels as the images at the previous level because of its greater coarseness. For such an image with a finite number of pixels, this wavelet representation is exact. Both of these figures are taken from the thesis of Mallat [15], in which he developed the wavelet basis as a tool for computer vision.

Because of the hierarchical nature of the wavelet basis, there is a simple accelerated transform method for this basis, which works in $O(N \log N)$ time for an expansion with N elements, as does the FFT. This makes the wavelet basis a practical tool as a general transform.

As a new application of wavelets, we propose to use them as a basis for numerical solution of differential equations, just as Fourier components are used in the spectral or pseudo-spectral method. The potential advantages of this method can be described in several ways. Because of the fast transform method for wavelets, the method would have the speed advantage of the spectral method. However because the wavelets are localized in space, it would still be possible to refine the representation at selected regions in which the solution has complicated behavior. For example in a problem with singularity development, such as the Schrodinger equation for laser beam propagation, large values of N corresponding to a refined mesh could be used near the singularity, while small values of N (coarse mesh) could be used away from the singularity. The main difficulties with implementing this suggestion is that derivatives do not act simply on the wavelet transform, as they do on the Fourier transform. This problem should be tractable, however.

A second potential advantage of the wavelet transform for scientific computations is its relation to the multi-grid method. In the multi-grid method, the solution is computed on a series of grids: on fine grids to eliminate high frequency errors and on coarse grids to eliminate low frequency errors. This method has proved quite successful in practice. The wavelet representation would represent these various

grids in a single transformation. At the same time it could allow the series of grids to differ from one region of space to another.

Parametric Estimation

In many signal processing problems, it is desirable to represent the data in terms of a parametrized set of functions [2,17]. This results in increased data compression and better physical intuition for the representation. For example in analysis of resonance spectroscopy data [17] it is desirable to represent the signal in terms of Lorentzians, with Fourier transforms of the form

$$\hat{r}(k) = (k - (k_r + k_i))$$

in which the real frequency k_r represents the frequency of the resonance and the imaginary frequency k_i represents the temporal decay.

There are several difficulties with such parametric representations. First the necessary number of elements is not known ahead of time and this freedom in the representation can cause extreme problems. For example, data corresponding to a single peak can easily be misinterpreted as corresponding to two peaks, or vice-versa. A second problem is that the optimal representation may not be easy to determine. This is partly for the first reason, but also because the parameterization is nonlinear. A third related difficulty is that in general there are no fast transform methods for these parametric estimation problems.

These are difficult problems which will require considerable effort, but the advantages of parametric estimation are sufficiently strong to justify it. As one possible approach, we suggest that the wavelet basis may be used as a first fast transform method. Then in a second stage the wavelet representation could be converted into the the desired parametric form. The second step would be less expensive than direct conversion of fourier data to parametric form, if the parametric form is localized in space or time.

3. Numerical Methods That Use Randomness

Random methods in scientific computation have proved successful in a variety of applications, including computation of large dimensional integrals, determination of statistical equilibria and gas dynamic simulations. Two recent successful random method are the random vortex method for viscous fluid flow [3] and the simulated annealing method for minimization [9,12,13]. However there is still some general reluctance to use random methods, because their validity is often difficult to assess and the random noise in the computational results is undesirable. Nevertheless, for many problems random

methods are the only ones which can provide reliable answers in realistic times, and for this reason they will continue to gain in use. For this reason, additional effort is needed to produce new random methods that are more accurate and to develop better methods for effectively analyzing the accuracy of such methods.

This report describes three random methods. The first two employ deterministic information to decrease the variance of a classical random method. The third is a successful recent method for optimization.

Stationary Phase Monte Carlo

Feynman path integrals have been greatly successful as a theoretical tool and, when modified to parabolic problems, they can be evaluated to give good numerical answers. However in their original form they are difficult to evaluate by the Monte Carlo method because of the large amount of oscillation that they contain. The desired integral is of the form

$$I = \int f(x) e^{i\phi(x)/\epsilon} dx$$

in which ϵ is quite small. In the limit of infinitesimally small ϵ the solution can be given by the method of stationary phase, which for physical problems amounts to the geometric optics solution. The Stationary Phase Monte Carlo method proposed by [6,14] attempts to use this information to make the evaluation of the integral I more feasible, even at intermediate values of ϵ .

The idea is to weight the integral in such a way that the paths of stationary phase, for which the oscillation is not strong, are sampled most often. While the other paths, for which canceling oscillations make their contribution much less significant, are sampled less frequently. As a result, the Monte Carlo integrals can be performed at a realistic cost in terms of number of trials performed.

A Hybrid Simulation Method for Rarefied Gas Dynamics

Flow of rarefied gases, as occurs in flight through the upper portions of the atmosphere or in dilute molecular beams, is described by the Boltzmann equation of kinetic theory. Because of the complexity of description (space, time and velocity are all independent variables) of this equation, its direct solution is quite difficult. The most successful numerical method at present is the Direct Simulation Monte Carlo method [1], in which the gas is represented by a set of particles with given velocities and positions. The particles move according to their velocities and undergo collisions which change their velocities.

Randomness is included in the initial choice of the particles and in the choice of collision outcomes. In this way, each particle can be thought of as representing a whole ensemble of particles.

In the small Knudsen number limit for which particles suffer many collisions in a short time, this method becomes quite expensive because of the need to compute the outcome of many collisions. However in this limit the molecular distribution function is well represented by a Maxwell distribution, since the large number of collisions quickly drives the distribution into a local thermodynamic equilibrium. In such a state the evolution of the gas is well described by the fluid dynamic equations (Euler or Navier-Stokes) with equation of state and transport coefficients that must be determined by some physical principle or by the Chapman-Enskog expansion. The need is for a method that will work well for intermediate Knudsen numbers, as well as such small Knudsen numbers. In many problems of spaceflight and transport theory, this disparity in the range of Knudsen numbers is one of the strongest difficulties.

As in the previous section, we propose to use the deterministic information, this time from fluid dynamics, in order to simplify the computational problem for the Boltzmann equation at intermediate Knudsen number. Represent the distribution function $F(\xi)$ as

$$F(\xi) = M(\xi) + f(\xi)$$

in which $M(\xi)$ is a maxwell distribution and $f(\xi)$ is a correction to M . One difficulty with this approach is that the correction term f need not be positive so that some particles should have a negative weight.

Effective use of a decomposition as above first requires solution of the corresponding linearized equation around the state M . In fact it seems that if the linearized problem can be effectively solved that the nonlinearities do not cause significant new difficulties.

Simulated Annealing

There are important unsolved problems regarding the random, heuristic search algorithm called simulated annealing. Annealing is the process of slowly cooling a physical system in order to obtain states with globally minimum energy. By simulating such a process, near globally-minimum-cost solutions can be found for very large optimization problems[12]. The applications include image restoration, combinatorial optimization (e.g. VLSI routing and placement), code design for communication systems and certain aspects of artificial intelligence. The theoretical tools for analysis include the theory of nonstationary Markov chains, statistical physics analysis techniques, large deviation theory and singular perturbation theory. The theory has been able to predict the large-time behavior of the annealing algorithm

(see [9] for a survey of the early theory). For example, Geman and Geman[8] first showed that if the temperature tends to zero as $C/\log(\text{Time})$ for a sufficiently large constant C , then the state of the annealing algorithm converges in probability to the set of states of globally minimum energy, and Hajek[10] showed that a necessary and sufficient condition on C is that C be at least as large as the depth of the deepest local minimum of the cost function. However, for practical applications, a much more useful thing to know about the algorithm is how its short-term performance behaves as a function of the size of the problem that the algorithm is applied to. For example, computer scientists know that maximum flows for graphs can be found in a number of computations that grows as the third power of the number of nodes and that there is no known algorithm for solving the traveling salesman problem in a number of computations that does not grow exponentially fast in the number of cities. There is very little known about how simulated annealing or other random algorithms fare in such a framework. One result in this direction was given by Sasaki and Hajek[16] who show that simulated annealing can find near-maximum matchings in a graph in average time that grows polynomially with the number of nodes in the graph, and they show the impossibility of finding the exact maximum matching in polynomial time when using simulated annealing. Further work in this direction might be based on the analysis of eigenvalues associated with large transition matrices--as indicated in the report of Jerrum and Sinclair[11], which generalizes N. Alon's bounds on eigenvalues based on certain mixing conditions.

4. Mathematics in control theory

Mathematics plays an important dual role in control theory according to the recent report of the Panel on Future Directions in Control Theory [7]. In the next three paragraphs we shall summarize some of the main points of the report. The first role of mathematics in control theory is to provide a structure in which control problems can be formulated in precise terms. The mathematical structure must be at once rich enough to describe the problems adequately, and simple enough to be mathematically tractable. The second role of mathematics in control theory is to bring the power of mathematical machinery to bear on the solution of control problems once they have been precisely formulated. Deep mathematical problems are frequently encountered, and answering them usually requires new mathematics. For example, most of the established mathematical results about partial differential equations which rely on continuity or other regularity properties of control inputs are often inadequate. The mathematical tools employed in control theory include complex function theory, linear and abstract algebra, ordinary and partial differential equations, variational methods, stochastic processes, functional analysis, operator theory, differential

geometry, Lie algebras, algebraic geometry, discrete mathematics, optimization and numerical analysis. The role that these mathematical theories play in control theory is often very different from that seen in traditional applied mathematics, so that in fact, control theory has established many unexpected connections between pure and applied mathematics. The report [7] describes many contemporary applications of control theory. Examples include applications to the space shuttle: algorithms to handle ascent and descent from orbit, studies of dynamic interactions between the shuttle and flexible payloads, optimal control to improve fuel economy of axis rotation maneuvers. Some other recent applications include a strategy to control the introduction of nonlinearly interacting agents in treatment of malignant brain tumors, a NASA helicopter autopilot control system, control of large power networks under adverse conditions, control systems in the F-16 and the experimental forward swept wing X-29, coordination and control of 36 mirrors in the Keck 10-meter telescope, damping in tall buildings, antilock brake systems, and many others. Finally, the report describes in some detail what modern science needs from control theory--there is an emphasis on large-scale control problems, nonlinear feedback, and adaptive or even self-organizing control.

5. References

1. Graeme Bird. "Monte Carlo simulation of gas flows," *Ann. Review Fluid Mech.* 10 (1978) 11-31.
2. Yoram Bresler and Albert Macovski. "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoustics, Speech and Signal Proc.* ASSP-34 (1986) 1081-1089.
3. Alexander Chorin. "The vortex method and boundary layer instability," *SIAM J. Sci. Stat. Comput.* 1 (1980) 1-21.
4. Ingrid Daubechies, A. Grossman and Y. Meyer. "Painless nonorthogonal expansions," *J. Math. Physics* 27 (1986) 1271-1283
5. Ingrid Daubechies. "Orthonormal bases of compactly supported wavelets," preprint (1988).
6. J.D. Doll, Thomas L. Beck and D.L. Freeman. "A quantum Monte Carlo dynamics: the stationary phase Monte Carlo path integral calculation of finite temperature time correlation functions," *J. Chem. Phys.* submitted.
7. Wendell H. Flemming. "Future directions in control theory--A mathematical prespective," *Society for Industrial and Applied Mathematics*, Philadelphia, 98 pp., 1988.

8. S. Geman and D. Geman "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 6, pp. 721-741, November 1984.
9. B. Hajek, "A tutorial survey of theory and applications of simulated annealing," Proc. IEEE Conf. on Decision and Control, December 1985.
10. B. Hajek, "Cooling schedules for optimal annealing," Mathematics of Operations Research, vol. 13, pp. 311-329, May 1988.
11. M. Jerrum and A. Sinclair, "Approximating the Permanent", Internal report CSR-275-88, Department of Computer Science, University of Edinburgh.
12. S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi "Optimization by simulated annealing," Science, vol. 220, pp. 621-680, May 1983.
13. Peter J.M. van Laarhoven and E.H.L. Aarts. Simulated Annealing: Theory and Application. Dordrecht, Boston: Reidel (1987).
14. Nancy Makri and William H. Miller, "Monte Carlo integration with oscillatory integrands: implications for Feynman path integration in real time," Chem. Phys. Letters 139 (1987) 10-14.
15. Stephan Mallat. "Multiresolution representations and wavelets," Thesis (1988) U. Penn.
16. G. Sasaki and B. Hajek, "The time complexity of maximum matching by simulated annealing," J. Assoc. for Computing Machinery, pp. 387-403, April 1988.
17. J. Tang and J.R. Norris. "LPZ Spectral analysis using linear prediction and the z-transform," J. Chem. Phys. 84 (1986) 5210-5211.

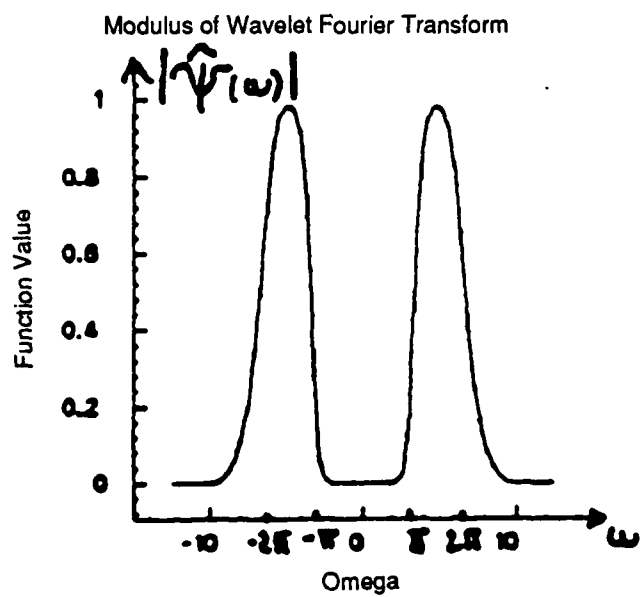
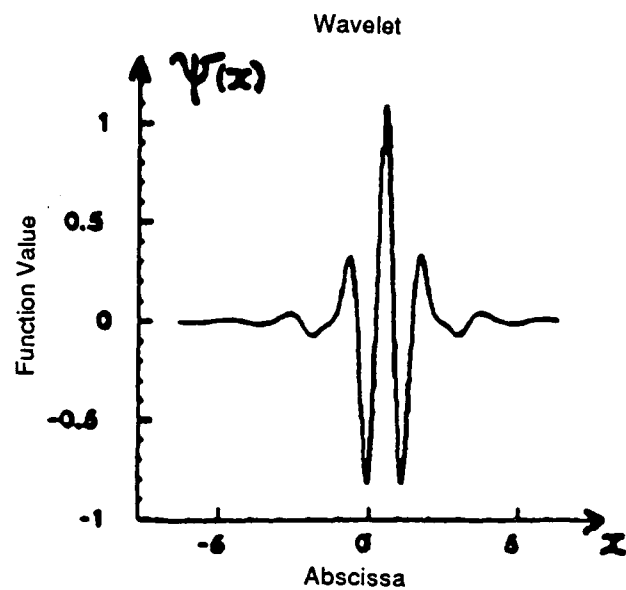
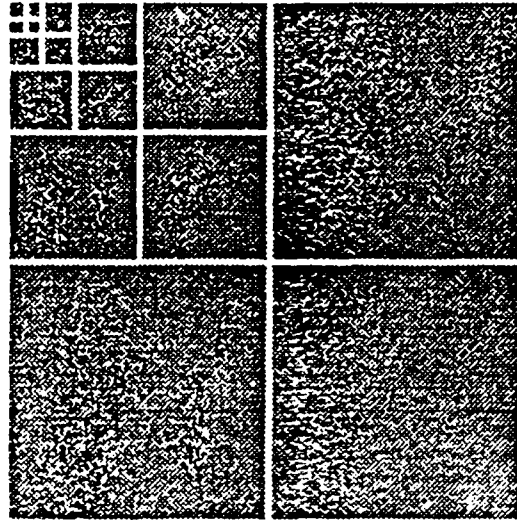


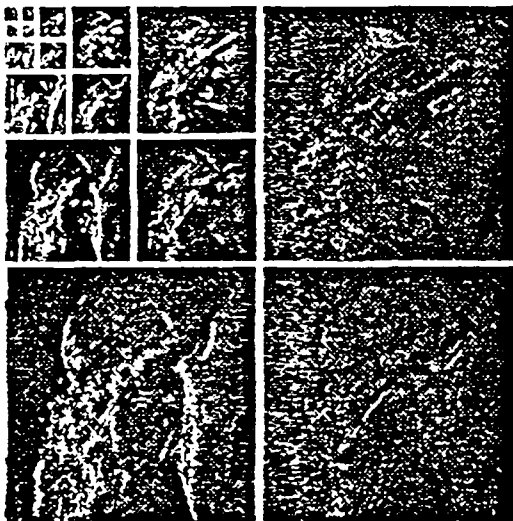
Figure 1.



(a)



(b)



(c)



(d)

(a): Original lady image. (b): Wavelet representation for $J = -5$. (c): Absolute value of the detail images pixels. (d): Reconstructed image from a representation with floating-point precision.

Figure 2.

IV. CONCLUSION

The Defense Science Study Group represents a long-term investment by the Department of Defense to raise the level of awareness of a new generation of this country's most outstanding young scientists and engineers to the scientific and technical issues of national security. As such, the success of this program will not be known for some time to come. In addition, even under the best of circumstances, the long-term results of the program offer no guarantee that future defense policy or political decisions dealing with national security issues will not destroy links that a program such as this seeks to create.

Nevertheless, investing in the purposeful creation of these links does have clear benefits in the short term and could prove to keep channels of communication open during times of stress in the long term. Some evidence of the short-term benefits is already apparent.

The first group of individuals going through the program has been influenced by DoD in a variety of expected and unexpected ways. A real interest in national defense problems developed among participants and this interest was communicated to colleagues and students. A result has been the guiding of some promising students toward careers in national defense areas. Some participants have modified existing research projects to reflect DoD interests as well as submitted and won grants and contracts from DoD for research in areas of interest.

Perhaps the most immediate impact is the growing involvement of DSSG members in the advisory apparatus of DoD. This has taken any of a number of forms. Some members serve as advisors to program managers in DoD. One has joined the JASON study group. Some have served on Defense Science Board studies and are under consideration to serve on other studies and task forces.

The interest, enthusiasm, and continued involvement of former participants in DoD activities is the best indication to date that the program appears to be succeeding in its objectives.

APPENDIX A

PARTICIPANTS AND MENTORS

PARTICIPANTS

RUSSEL CAFLISCH	Courant Institute of Mathematical Sciences New York University
STEVEN K. CASE	Electrical Engineering University of Minnesota
VICKI L. CHANDLER*	Molecular Biology University of Oregon
KATHERINE T. FABER	Ceramic Engineering Northwestern University
BRUCE HAJEK*	Electrical and Computer Engineering University of Illinois
DEBORAH A. JOSEPH*	Computer Science University of Wisconsin
RANDY H. KATZ*	Computer Science University of California, Berkeley
STEVEN E. KOONIN	Theoretical Physics California Institute of Technology
FREDERICK K. LAMB	Physics University of Illinois
NATHAN S. LEWIS	Chemistry California Institute of Technology
PHILIP S. MARCUS	Astronomy and Applied Mathematics University of California, Berkeley

THOMAS F. ROSENBAUM	Physics, James Franck Institute University of Chicago
STEVEN J. SIBENER	Chemistry, James Franck Institute University of Chicago
DANIEL J. STEIN	Physics University of Arizona
WARREN S. WARREN	Chemistry Princeton University
R. STANLEY WILLIAMS*	Chemistry University of California, Los Angeles
W. HUGH WOODIN	Mathematics University of California, Berkeley

* Added in 1987 and will continue as a member in 1989.

MENTORS

LEW ALLEN, JR.*	Director, Jet Propulsion Laboratory General, U.S. Air Force (Retired)
DANIEL ALPERT	Director, Program in Science, Technology and Society University of Illinois
RICHARD BERNSTEIN	Professor of Chemistry University of California, Los Angeles
R. STEPHEN BERRY	Professor of Chemistry University of Chicago
SOLOMON J. BUCHSBAUM	Executive Vice President, Customer Systems AT&T Bell Laboratories
PETER CARRUTHERS	Head, Department of Physics University of Arizona
RUTH DAVIS**	President The Pymatuning Group, Inc.
ALEXANDER H. FLAX	Private Consultant
EUGENE G. FUBINI*	Private Consultant
ANDREW J. GOODPASTER	General, U.S. Army (Retired) Private Consultant
ISAAC C. KIDD, JR.**	Admiral, U.S. Navy (Retired) Private Consultant
MARTHA KREBS	Associate Director, Planning & Development Lawrence Berkeley Laboratory

PETER LAX

Director, Courant Institute
New York University

S. SOLOMON PENNER

Director, Center for Energy and
Combustion Research, University of California, San
Diego

DAVID PINES

Professor of Physics and Electrical Engineering
University of Illinois

WILLIAM H. PRESS

Chairman, Department of Astronomy
Harvard University

HERBERT YORK

Director, Science, Technology and
Public Affairs, University of California, San Diego

* Left in 1987

** Added in 1987

APPENDIX B

SOME DISTINGUISHED SPEAKERS

"

Anthony Battista	Staff Member, House Armed Services Committee, U.S. Congress
Antonia Chayes	Chairman, Endispute, Inc., and Former Under Secretary of the Air Force
Robert Cooper	President, Pollard Road, Inc., and Former Director, DARPA
Charles Fowler	Chairman, Defense Science Board
John Grimes	Director of National Security Telecommunications and Director of Defense Programs (C ³), National Security Council
Robert Hillyer	Technical Director, Naval Ocean Systems Center
Robert Howard	Deputy Chief, National Security Division, Office of Management and Budget
VADM David Jeremiah	Director, Navy Program Planning Office
Donald Latham	Assistant Secretary of Defense for Command, Control, Communications and Intelligence (C ³ I), DoD
VADM William Ramsey	Deputy Commander-in-Chief, U.S. Space Command
Victor Reis	Corporate Vice President, Science Applications International (SAI)
Jack Ruina	Director, Defense and Arms Control Studies Program, MIT, and Former Director, DARPA
Philip Selwyn	Director, Office of Naval Technology

AD NUMBER

E501188 *****

FIELD 2:	FLD/GRP(S)
FIELD 3:	ENTRY CLASS.
FIELD 4:	NTIS PRICE
FIELD 5:	SOURCE NAME
FIELD 6:	UNCLASS. TITLE
FIELD 7:	CLASS. TITLE
FIELD 8:	TITLE CLASS.
FIELD 9:	DESCRIPTIVE NOTE
FIELD 10:	PERSONAL AUTHORS
FIELD 11:	REPORT DATE
FIELD 12:	PAGINATION
FIELD 13:	PROCESSING LEVEL
FIELD 14:	REPORT NUMBER
FIELD 15:	CONTRACT NUMBER
FIELD 16:	PROJECT NUMBER
FIELD 17:	TASK NUMBER
FIELD 18:	MONITOR ACRONYM
FIELD 19:	MONITOR SERIES
FIELD 20:	REPORT CLASS
FIELD 21:	SUPPLEMENTARY NOTE
FIELD 22:	ALPHA LIMITATIONS
FIELD 23:	DESCRIPTORS
FIELD 24:	DESCRIPTOR CLASS.
FIELD 25:	IDENTIFIERS
FIELD 26:	IDENTIFIER CLASS.
FIELD 27:	ABSTRACT
FIELD 28:	ABSTRACT CLASS.
FIELD 29:	INITIAL INVENTORY
FIELD 30:	ANNOTATION
FIELD 31:	SPECIAL INDICATOR
FIELD 32:	REGRADE CATEGORY
FIELD 33:	LIMITATION CODES
FIELD 34:	SOURCE SERIAL
FIELD 35:	SOURCE CODE
FIELD 36:	DOCUMENT LOCATION
FIELD 37:	CLASSIFIED BY
FIELD 38:	DECLASSIFY ON
FIELD 39:	DOWNGRADE TO CONF OF
FIELD 40:	GEOPOLITICAL CODE
FIELD 41:	TYPE CODE

150300 U

[[INSTITUTE FOR DEFENSE ANALYSES ALEXANDRIA VA
[3]UNMARY [REPORT OF THE [DEFENSE [SCIENCE [STUDY [GROUP, 1985-1988

U
[F]INAL REPT. [O]CT 85-[N]OV 88.
[B]ERGMANN, [R]ICHARD [J].: [L]ICATO, [N]ANCY [P].
SEP 89
148P

[IDA]-[P]-2310
[MDA]903-84-[C]-0031

**U
[IDA/HQ], [SBI
89-34853, [AD]-[E]501 186**

~~[A]VAILABILITY CONTROLLED BY [IDA], [ATTN: [FIS], [A]LEXANDRIA, [VA-122311,
[A]NAGEMENT ONLY; DOCUMENT WILL BE MADE AVAILABLE FROM [DTIC] AFTER PROCESSING.
*[DEFENSE SYSTEMS], [RESEARCH MANAGEMENT] [MILITARY RESEARCH], [DEPARTMENT OF DEFENSE],
[ARMS CONTROL], [NATIONAL DEFENSE], [SCIENTISTS], [THREAT EVALUATION], [USSR].~~

[LPN]-[IDA]-[A]-103. [SBI]1. [FISCAL YEAR 1990. [D]EFENSE [S]CIENCE [S]TUDY [G]ROUP.

U [T]HE [D]EFENSE [S]CIENCE [S]TUDY [G]ROUP IS A [DARPA] SPONSORED PROGRAM THAT IDENTIFIES SOME OF THE MOST TALENTED YOUNG SCIENTISTS AND ENGINEERS IN THE COUNTRY TODAY AND EXPOSES THEM TO THE OUTSTANDING ISSUES OF NATIONAL DEFENSE. [T]HE PROGRAM SEEKS TO STRENGTHEN TIES WITH THE PART OF THE SCIENTIFIC AND ENGINEERING COMMUNITY THAT IS OUTSIDE OF THE DEFENSE ESTABLISHMENT. [I]TS GOALS ARE TO PROVIDE AN EDUCATION TO THIS SELECT GROUP OF INDIVIDUALS ON THE BROAD SCOPE OF CRITICAL DEFENSE-RELATED SCIENTIFIC AND TECHNICAL PROBLEMS, ENCOURAGE LONG-TERM, PERSONAL INVOLVEMENT IN THEIR SOLUTION, AND PROVIDE [D]O[D] WITH A NEW SOURCE OF TECHNICAL ADVISORS AND INFORMED CRITICS. [T]HE PROGRAM RESIDES AT [IDA] AND IS GUIDED BY THE ADVICE OF MENTORS WHO HAVE HAD DISTINGUISHED CAREERS IN DEFENSE OR ACADEMIA. [I]NDIVIDUALS SPEND ABOUT TWENTY DAYS A YEAR FOR THREE YEARS LISTENING TO PRESENTATIONS GIVEN BY DISTINGUISHED SPEAKERS. VISITING DEFENSE FACILITIES AND CONDUCTING DEFENSE STUDIES ON TOPICS OF INTEREST. [T]HE REPORT SUMMARIZES THE PROGRAM'S FIRST THREE YEARS OF ACTIVITIES FROM 1985-1988.

 ϕ

1 ET
[F 179350

5108

FIELD 42: IAC ACCESSION NO.
FIELD 43: IAC DOCUMENT TYPE
FIELD 44: IAC SUBJECT TERM
FIELD 45: EXTENDED BY
FIELD 46: REVIEW ON DATE
FIELD 47: REASON CODE
FIELD 48: SBIE SITE SYMBOLS

eeeeee

IDAHO34853